

Una chiacchierata sulla logica probabilistica

L'idea che la logica possa essere integrata con la probabilità non nasce come un semplice tentativo di “ammorbidire” la verità classica, sostituendo il valore 1 con un numero reale compreso tra 0 e 1. La tesi più profonda, che attraversa tutta la storia delle logiche probabilistiche, è che la probabilità riveli una struttura già implicita nella semantica logica: quando ragioniamo su insiemi di modelli, su insiemi di interpretazioni, su insiemi di mondi possibili, stiamo già maneggiando oggetti che possono essere misurati, pesati, confrontati. La probabilità non è un'aggiunta esterna: è un modo naturale di quantificare la “prevalenza” delle strutture che soddisfano una formula.

L'idea di associare una misura di probabilità alla semantica logica nasce dal riconoscimento che l'insieme dei modelli di un linguaggio formale non è soltanto un dominio qualitativo, ma può essere trattato come uno spazio misurabile. Consideriamo un linguaggio del primo ordine L con segnatura $\sigma = (C, F, R)$, dove C è l'insieme delle costanti, F dei simboli di funzione e R dei simboli di relazione. L'insieme Q di tutte le strutture $M = (D, I)$ che interpretano L costituisce lo spazio dei mondi possibili. Ogni formula φ del linguaggio definisce un sottoinsieme di Q , cioè l'insieme $[\varphi] = \{ M \in Q \mid M \models \varphi \}$. L'idea fondamentale è dotare Q di una σ -algebra Σ tale che ogni insieme $[\varphi]$ sia misurabile, e definire una misura $\mu : \Sigma \rightarrow [0,1]$ che rispetti la struttura logica. In particolare, si richiede che $\mu([\varphi]) = 1$ se φ è una tautologia, che $\mu([\varphi \vee \psi]) = \mu([\varphi]) + \mu([\psi]) - \mu([\varphi \wedge \psi])$, e che i quantificatori siano interpretabili come limiti di frequenze: $\mu([\forall x \varphi(x)]) = \lim_{n \rightarrow \infty} \mu([\varphi(c_1) \wedge \dots \wedge \varphi(c_n)])$, dove $c_1, c_2, \dots$ sono nuove costanti che percorrono il dominio. Questa condizione, apparentemente artificiale, cattura l'idea che il quantificatore universale esprima una proprietà che vale “quasi ovunque” nel dominio, e che la misura su Q rifletta la frequenza con cui le interpretazioni delle costanti soddisfano la formula.

Consideriamo un linguaggio del primo ordine L , con firma $\sigma = (C, F, R)$ composta da insiemi numerabili di simboli di costante, di funzione e di relazione. Una struttura M per L è una coppia (D^M, I^M) , dove D^M è un dominio non vuoto e I^M interpreta ogni simbolo: a ciascuna costante $c \in C$ associa un elemento $c^M \in D^M$, a ciascun simbolo di funzione n -aria $f \in F$ associa una funzione $f^M : (D^M)^n \rightarrow D^M$, e a ciascun simbolo di relazione n -aria $R \in R$ associa un sottoinsieme $R^M \subseteq (D^M)^n$. Indichiamo con Q l'insieme di tutte le strutture per L (a isomorfismo trascurato, o più precisamente scegliendo un insieme rappresentativo di strutture, se vogliamo evitare problemi di taglia).

A ogni formula chiusa φ di L (cioè senza variabili libere) associamo il sottoinsieme di Q definito da

$$[\varphi] = \{ M \in Q \mid M \models \varphi \}.$$

L'idea di Gaifman è di considerare la collezione di tutti questi insiemi $[\varphi]$ come generatori di un'algebra (o σ -algebra) di sottoinsiemi di Q , e di definire su di essa una misura di probabilità μ tale che la combinatoria logica dei connettivi e dei quantificatori si rifletta esattamente nelle proprietà della misura. Il primo passo è considerare l'algebra booleana B generata da tutti gli insiemi $[\varphi]$. Se L è numerabile, l'insieme delle formule è numerabile, e quindi B è una algebra booleana numerabile generata da un insieme numerabile di elementi. In molti trattamenti si lavora direttamente sull'algebra di Lindenbaum-Tarski di L : si considera l'insieme delle formule modulo equivalenza logica ($\varphi \equiv \psi$ se e solo se $\models \forall (\varphi \leftrightarrow \psi)$), e lo si quozienta rispetto all'ideale delle contraddizioni. L'algebra risultante, che possiamo denotare $LT(L)$, è una algebra booleana in cui ogni classe $[\varphi]$ corrisponde a quella che, a livello insiemistico, era l'insieme dei modelli di φ .

Consideriamo ora una funzione $\mu: B \rightarrow [0,1]$. Vorremmo che μ soddisfacesse le seguenti condizioni:

1. $\mu(Q) = 1$, cioè $\mu(\top) = 1$ (le tautologie hanno probabilità 1, gli insiemi universalmente validi hanno misura 1);
2. $\mu(\emptyset) = 0$, cioè $\mu(\perp) = 0$;
3. se $A, B \in B$ sono disgiunti ($A \cap B = \emptyset$), allora $\mu(A \cup B) = \mu(A) + \mu(B)$;
4. più in generale, per ogni famiglia numerabile $(A_i)_i$ di insiemi mutualmente disgiunti in B , $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$.

La condizione 4) è la σ -additività. Se la imponiamo solo per famiglie finite, parliamo di misura finitamente additiva; nel contesto di Gaifman si guarda fin da subito alla σ -additività, ma una parte importante della tecnica consiste nel costruire prima misure finitamente additive e poi passare alla σ -additività tramite completamenti e teoremi di estensione del tipo Carathéodory.

Nel quadro specifico di Gaifman, la costruzione di μ è guidata da una condizione supplementare che riguarda il comportamento dei quantificatori. La notazione $[\exists x \ \phi(x)]$ indica l'insieme delle strutture in cui esiste almeno un elemento del dominio che soddisfa ϕ . Per ragioni tecniche, è conveniente lavorare con nuovi simboli di costante introdotti per “attualizzare” vari elementi del dominio. Se $(c_i)_i$ è una sequenza infinita di nuove costanti che non compaiono in L , si considerano formule come $\phi(c_1)$, $\phi(c_2)$, ..., $\phi(c_n)$, e le relative congiunzioni finite. L'idea è che la formula $\forall x \ \phi(x)$ debba essere riflessa nella misura come limite di una condizione su tutte le costanti: “la probabilità che ogni costante c_i soddisfi ϕ tende a 1”. Formalmente, una condizione tipica che si impone è di questo tipo: per una formula $\psi(x)$ a una variabile e una sequenza $(c_i)_i$ di nuove costanti, la probabilità di $\forall x \ \psi(x)$ è il limite della probabilità di $\bigwedge_{i=1}^n \psi(c_i)$ quando $n \rightarrow \infty$:

$$\mu([\forall x \ \psi(x)]) = \lim_{n \rightarrow \infty} \mu([\bigwedge_{i=1}^n \psi(c_i)]).$$

Occorre essere precisi: il significato rigoroso di quest'uguaglianza dipende da come costruiamo la famiglia di costanti e da come estendiamo il linguaggio. In alcune presentazioni, i limiti vengono interpretati come $\limsup$ o $\liminf$, e si introducono assiomi che garantiscono la consistenza di tali valori. Il punto centrale è che l'estensione da formule senza quantificatori a formule con quantificatori avviene passando da eventi che riguardano un numero finito di costanti a eventi che combinano asserzioni su un infinito numerabile di costanti.

L'enunciato fondamentale, in un formato tipico, è il seguente.

Teorema (Gaifman, esistenza di misure logiche su L). Sia L un linguaggio del primo ordine numerabile. Esiste una misura di probabilità μ definita sulla σ -algebra generata dagli insiemi $[\phi]$, con ϕ formula chiusa di L , tale che:

- (i) per ogni tautologia ϕ (cioè formula valida in tutti i modelli di L), $\mu([\phi]) = 1$;
- (ii) per ogni coppia di formule chiuse ϕ, ψ , vale $\mu([\phi \vee \psi]) = \mu([\phi]) + \mu([\psi]) - \mu([\phi \wedge \psi])$;
- (iii) per ogni formula $\psi(x)$ e per ogni sequenza di nuove costanti $(c_i)_i$, si ha

$$\mu([\forall x \ \psi(x)]) = \lim_{n \rightarrow \infty} \mu([\bigwedge_{i=1}^n \psi(c_i)]),$$

dove i limiti sono ben definiti e compatibili con la struttura di misura.

Dal punto di vista logico, questo significa che possiamo interpretare la probabilità come misura sui modelli senza violare la struttura logica di L : le congiunzioni diventano intersezioni, le disgiunzioni unioni, le negazioni complementi, e i quantificatori si collegano a limiti di eventi che coinvolgono sequenze di costanti.

Una dimostrazione tipica procede per approssimazioni finite e uso di ultrapotenze. L'idea è costruire una famiglia di modelli finiti per sottolinguaggi finiti di L , dotati di una distribuzione di probabilità uniforme (o comunque ben controllata), e poi passare a un limite induttivo (nel senso categorico) o a un ultralimite (nel senso dei filtri ultrafiniti) che produce un modello “grande” in cui le frequenze si stabilizzano e la misura si estende coerentemente.

Più concretamente, si procede così. Poiché L è numerabile, possiamo enumerare le sue formule chiuse come $(\varphi_n)_n$. Consideriamo una struttura combinatoria che approssima i possibili valori di verità di queste formule come se fossero “coordinate” di un punto in spazio prodotto $\{0,1\}^{\mathbb{N}}$. Ogni punto $\omega \in \{0,1\}^{\mathbb{N}}$ corrisponde a un possibile assegnamento di verità: $\omega(n) = 1$ se φ_n è “vera” in ω , $\omega(n) = 0$ altrimenti. Lo spazio di tutte queste sequenze è una specie di “Stone space grezzo” associato al linguaggio. Perché questo abbia senso logico, dobbiamo restringerci a quelle sequenze che rispettano le relazioni derivanti dalle tautologie e dalle conseguenze logiche; in altre parole, a quelle sequenze che corrispondono agli ultrafiltri sull'algebra di Lindenbaum-Tarski. Denotiamo con $S(L)$ l'insieme di tutti questi ultrafiltri: ogni elemento $U \in S(L)$ è un insieme massimale di formule chiuse consistente, equivalente a una teoria completa. È un fatto standard che $S(L)$ può essere visto come uno spazio di Stone compatto e totalmente sconnesso, dotato di una topologia generata dagli insiemi della forma

$$[\varphi]^S = \{ U \in S(L) \mid \varphi \in U \}.$$

Questa collezione di insiemi clopen coincide, a livello insiemistico, con la “traccia” di $[\varphi]$ sull'insieme delle teorie complete.

La costruzione di μ può essere condotta su $S(L)$: si definisce una misura di Borel μ_S su $S(L)$ tale che $\mu_S([\varphi]^S) = \mu([\varphi])$ per tutte le formule chiuse φ . In questa formulazione, il teorema di Gaifman diventa: esiste una misura di Borel probabilistica μ_S su $S(L)$ per la quale ogni teoria consistente E (cioè ogni insieme di formule chiuse senza contraddizione) è contenuta in un insieme di misura 1:

$$\mu_S(\bigcap_{\varphi \in E} [\varphi]^S) = 1.$$

In altri termini, per ogni E consistente esiste una misura che “concentra” la probabilità sulle teorie che estendono E .

Per costruire μ_S , si procede come segue. Si considera un sistema diretto di strutture finite M_i per sottolinguaggi finiti $L_i \subseteq L$. Per ogni M_i si definisce una misura uniforme sui sottoinsiemi del dominio e sulle interpretazioni, in modo che per ognuna delle formule chiuse di L_i si possa definire la frequenza di verità rispetto a M_i . Se M_i ha dominio di cardinalità N_i , la misura uniforme su $D^{\{M_i\}}$ è semplicemente la misura di conteggio normalizzata: per ogni $A \subseteq D^{\{M_i\}}$, $\mu_i(A) = |A| / |D^{\{M_i\}}|$. A questo punto, ogni struttura M_i induce una valutazione di verità per le formule di L_i , e dunque una distribuzione di frequenza sui valori di verità delle φ_n che appartengono a L_i . Il passaggio al limite usa un ultrafiltro non principale U su $\mathbb{N}$: si considerano le sequenze $(M_i)_i$ e si costruisce l'ultrapotenza $\prod_i M_i / U$. In questa ultrapotenza, i limiti delle frequenze si stabilizzano: se per ogni formula chiusa φ si definisce la frequenza relativa di M_i aventi φ vera, il limite rispetto all'ultrafiltro è ben definito e genera una funzione μ_∞ sulle formule.

Il punto tecnico cruciale è che μ_∞ così definita è finitamente additiva e soddisfa le condizioni che riflettono i connettivi logici. Per ottenere la σ -additività, si invoca un argomento alla Carathéodory: si estende μ_∞ dall'algebra finitamente generata (o dall'algebra cilindrica) alla σ -algebra generata, mantenendo l'additività numerabile. La coerenza con i quantificatori deriva dal modo in cui le sequenze di costanti vengono interpretate nelle ultrapotenze: le costanti aggiunte generano un sottoinsieme “tipico” del dominio e le frequenze delle formule $\psi(c_i)$ convergono alla misura dell'insieme $\{a \in D \mid M \models \psi(a)\}$. Dietro questo c'è la combinazione di due fatti: il lemma di Łoś (che assicura la conservazione delle verità in ultrapotenze) e la proprietà che, nella costruzione, la scelta dei M_i e la loro crescita in cardinalità sono tali da garantire che l'effetto delle costanti nuove “si diluisca” rispetto al dominio.

Una volta accettata la legittimità di questa costruzione, possiamo interpretare $\mu([\phi])$ come “probabilità che una struttura scelta a caso da Q (secondo μ) soddisfi ϕ ”. Il quadro formale è ora sufficientemente robusto da accogliere anche la prospettiva epistemica di Carnap. Carnap, tuttavia, sceglie un'impostazione più restrittiva e più vicina alle applicazioni induttive: invece di lavorare con l'intero L , si focalizza su linguaggi monadici con un numero finito di predicati e individui, e considera funzioni di conferma $c(h,e)$ che rappresentano il grado di conferma che un'evidenza e conferisce a un'ipotesi h .

Nel caso più semplice, sia L un linguaggio con un solo predicato monadico Q e individuali $a_1, \dots, a_n$ osservati. Supponiamo di aver osservato che s di questi n individui godono di Q , mentre $n-s$ non lo fanno. L'ipotesi h è che il prossimo individuo $a_{\{n+1\}}$ abbia proprietà Q , cioè h è la formula $Q(a_{\{n+1\}})$. La funzione di conferma di Carnap della sua famiglia λ -continua è definita da

$$c_\lambda(h \mid e_n) = (s + \lambda / (2K)) / (n + \lambda),$$

dove K è il numero di predicati nel linguaggio (qui $K=1$) e $\lambda \geq 0$ è un parametro di “forza a priori”. Il senso di questa formula è chiaro: al numeratore abbiamo il numero di successi osservati s più un termine a priori $\lambda / (2K)$ che rappresenta una inclinazione iniziale a credere che Q valga con probabilità $1/2K$ (per motivi simmetrici), e al denominatore il totale degli individui osservati n più λ , che rappresenta una “quantità fittizia” di osservazioni a priori.

Il risultato fondamentale che inquadra questa famiglia all'interno della teoria bayesiana è il seguente. Consideriamo il modello binomiale: supponiamo che ogni individuo abbia proprietà Q con probabilità θ , indipendentemente dagli altri, dove $\theta \in [0,1]$ è un parametro sconosciuto. La probabilità di osservare s successi in n osservazioni, condizionata a θ , è

$$\Pr(D_n \mid \theta) = \theta^s (1-\theta)^{n-s},$$

dove D_n è il dato osservato. Una funzione di conferma $c(h|e)$ è detta regolare e simmetrica se soddisfa una serie di condizioni (non dogmatismo, continuità, invarianza rispetto a permutazioni degli individui, ecc.) che tecnicamente garantiscono che essa possa essere vista come una probabilità condizionata rispetto a una misura a priori π su θ . In particolare, si richiede che ci sia una rappresentazione del tipo

$$c(h \mid e) = \int_0^1 f(h, \theta) \Pr(e \mid \theta) d\pi(\theta) / \int_0^1 \Pr(e \mid \theta) d\pi(\theta),$$

dove $f(h, \theta)$ è la probabilità dell'ipotesi h dato θ (nel caso di $h = Q(a_{\{n+1\}})$, $f(h, \theta) = \theta$).

Il teorema (attribuito a Carnap e Stegmüller in varie forme) afferma che, sotto ipotesi ragionevoli di regolarità e simmetria, ogni funzione di conferma c può essere rappresentata in questo modo per una

qualche misura a priori π su $[0,1]$. Nel caso specifico della famiglia c_λ , la misura π è una distribuzione Beta con parametri dipendenti da λ e K . Più precisamente, se prendiamo π come $\text{Beta}(\alpha, \beta)$ con $\alpha = \lambda / (2K)$, $\beta = \lambda / (2K)$, la probabilità a posteriori di successo per il prossimo individuo dopo aver osservato s successi su n prove è

$$E[\theta | D_n] = (s + \alpha) / (n + \alpha + \beta) = (s + \lambda/(2K)) / (n + \lambda),$$

che coincide con la definizione di $c_\lambda(h|e_n)$. La dimostrazione di questo fatto è un esercizio standard di calcolo bayesiano: si moltiplica la densità a priori $\theta^{\alpha-1} (1-\theta)^{\beta-1}$ per la verosimiglianza $\theta^s (1-\theta)^{n-s}$, si ottiene una densità a posteriori proporzionale a $\theta^{s+\alpha-1} (1-\theta)^{n-s+\beta-1}$, cioè ancora una Beta con parametri aggiornati, e si calcola il valore atteso di θ sotto questa distribuzione. Il punto concettuale è che l'intera famiglia di Carnap può essere vista come un caso particolare di un processo bayesiano parametricamente specificato; inoltre, il teorema rappresentazionale mostra che, in un senso preciso, non stiamo perdendo generalità: le funzioni di conferma che vogliamo considerare "razionali" nel senso logico-carnapiano sono esattamente quelle che si possono rappresentare come integrali rispetto a una misura a priori su un parametro.

Questa connessione chiude il cerchio con la costruzione misura-teoretica di Gaifman: anche lì, in fondo, si tratta di definire una misura su uno spazio di modelli (equivalente a uno spazio di parametri in un senso più ampio) e di interpretare le probabilità di formule come integrali di funzioni caratteristiche rispetto a questa misura. Dove Carnap lavora con uno spazio parametricamente semplice $[0,1]$ e con funzioni di verosimiglianza binomiali, Gaifman opera in uno spazio di modelli molto più complesso; ma il principio è analogo: la probabilità logica è una media rispetto a uno spazio sottostante.

Lo schema di Scott e Krauss porta questa idea a un livello di generalità ancora più alto, posizionando la misura direttamente sull'algebra di Lindenbaum-Tarski. Partiamo dall'algebra booleana $LT(L)$, formata da tutte le classi di equivalenza $[\varphi]$ di formule chiuse modulo equivalenza logica. Una funzione $\mu: LT(L) \rightarrow [0,1]$ è detta misura logica se soddisfa:

1. $\mu([T]) = 1$, $\mu([\perp]) = 0$;
2. se $[\varphi] \wedge [\psi] = [\perp]$ (cioè le formule φ, ψ sono logicamente incompatibili, nel senso che $\models \neg(\varphi \wedge \psi)$), allora $\mu([\varphi] \vee [\psi]) = \mu([\varphi]) + \mu([\psi])$;
3. se $([\varphi_i])_i$ è una famiglia numerabile di elementi mutuamente disgiunti di $LT(L)$ (cioè $\varphi_i \wedge \varphi_j$ è contraddittoria per $i \neq j$), allora

$$\mu(\vee_i [\varphi_i]) = \sum_i \mu([\varphi_i]).$$

La domanda è: esiste sempre uno spazio di probabilità (Ω, Σ, P) e, per ogni formula φ , una funzione misurabile $f_\varphi: \Omega \rightarrow \{0,1\}$ tale che

$$\mu([\varphi]) = P(\{\omega \in \Omega \mid f_\varphi(\omega) = 1\})?$$

E, inoltre, questo spazio è unico a meno di isomorfismo di σ -algebre di misura?

Il teorema di rappresentazione di Scott risponde affermativamente: ogni misura logica finitamente additiva e σ -additiva su $LT(L)$ ammette una rappresentazione di questo tipo. La dimostrazione si basa sull'idea che $LT(L)$, quotata rispetto all'ideale degli insiemi di misura zero, formi un'algebra di misura astratta (un "measure algebra") che, per un noto teorema di rappresentazione, può sempre essere vista come l'algebra di insiemi misurabili modulo nulli di qualche spazio di probabilità classico. Più concretamente, si prende la σ -algebra Σ_0 generata dagli insiemi atomici corrispondenti

ai valori di verità di un sistema di formule che separa i punti, e si definisce Ω come lo spazio di ultrafiltri compatibili con μ (in qualche presentazione, Ω coincide con $S(L)$, in altre è una costruzione più tecnica). Su Σ_0 si definisce P in modo che $P([\varphi]) = \mu([\varphi])$; poi si estende P alla σ -algebra generata, ottenendo una misura completa. Le funzioni f_φ non sono altro che le funzioni caratteristiche degli insiemi di Ω in cui φ è “vera”: $f_\varphi(\omega) = 1$ se ω contiene $[\varphi]$, 0 altrimenti.

L’unicità fino a isomorfismo di measure algebras è un risultato generale: se $(\Omega_1, \Sigma_1, P_1)$ e $(\Omega_2, \Sigma_2, P_2)$ sono spazi di probabilità tali che le loro algebre delle classi di equivalenza modulo insiemi di misura zero sono isomorfe come algebre di misura, allora c’è un isomorfismo (nel senso di misura-preservazione) tra i due spazi che rispetta l’identificazione degli insiemi misurabili. Applicato qui, significa che, se esistono due rappresentazioni per μ , esse sono essenzialmente la stessa, salvo rinominare gli elementi di Ω .

In questo modo, la misura logica μ non è un oggetto misterioso “interno” all’algebra delle formule: è equivalente a una misura classica su uno spazio di probabilità, e ogni formula viene interpretata come una variabile casuale $\{0,1\}$ -valuta. La probabilità logica di φ è la probabilità, nel senso usuale, che la variabile f_φ valga 1.

Queste tre linee – Gaifman, Carnap e Scott-Krauss – sono tre incarnazioni della stessa intuizione: la logica può essere immersa nella teoria della misura, e la probabilità logica non è un’aggiunta decorativa, ma una scelta di misura su uno spazio strutturato da L . In un passo successivo, questa struttura viene arricchita da quantificatori probabilistici (Keisler) e da nozioni di scambiabilità e rappresentazione come miscele di misure prodotto (Hoover), ma queste richiedono già di lavorare sistematicamente con domini misurati e con misure su insiemi di strutture piuttosto che su singole algebre di formule.

L’idea di Keisler nasce dall’osservazione che, se prendiamo sul serio la lettura “frequenziale” dei quantificatori del primo ordine, questi possono essere visti come affermazioni sulla misura (o sulla probabilità) dell’insieme degli elementi che soddisfano una certa proprietà. Il quantificatore universale $\forall x \varphi(x)$ dice, nella lettura classica, che tutti gli elementi del dominio soddisfano φ ; in una lettura misura-teoretica si potrebbe dire che l’insieme $\{ a \in D \mid M \models \varphi(a) \}$ ha misura 1. Analogamente, il quantificatore esistenziale $\exists x \varphi(x)$ può essere visto come “l’insieme di soluzioni ha misura positiva”. Keisler raffina questa intuizione introducendo quantificatori che esprimono direttamente vincoli di misura: il costrutto $(P x \geq r) \varphi(x)$ significa “la misura dell’insieme di x tali che $\varphi(x)$ è soddisfatta è almeno r ”, dove r è un numero reale in $[0,1]$.

Per rendere rigorosa questa estensione, si parte da un linguaggio del primo ordine L , si aggiunge una famiglia di nuovi quantificatori che hanno la forma $(P x \geq r)$ per ogni r razionale in $[0,1]$, e si definisce la sintassi ricorsivamente: se φ è una formula e x è una variabile, allora $(P x \geq r) \varphi$ è una formula. La semantica richiede che ogni struttura M sia arricchita da una misura definita sul dominio: formalmente, una struttura per questa logica è una terna $M = (D, I, \mu_D)$, dove D è il dominio non vuoto, I è l’interpretazione dei simboli come nel primo ordine classico, e μ_D è una misura finitamente additiva definita su una σ -algebra (o almeno su un’anello) di sottoinsiemi di D , contenente tutti gli insiemi definibili del tipo $\{ a \in D \mid M \models \varphi(a) \}$ per formule φ con parametri presenti in M . La definizione del soddisfacimento è la seguente: per una formula $\varphi(x)$ con una variabile libera x e un numero razionale r , si pone

$$M \models (P x \geq r) \varphi(x) \text{ se e solo se } \mu_D(\{ a \in D \mid M \models \varphi(a) \}) \geq r.$$

Il problema è assicurarsi che questa semantica sia sensata e che esista un sistema assiomatico completo per questa logica. Keisler propone un insieme di assiomi e regole che estendono quelli della

logica del primo ordine con schemi specifici per i quantificatori probabilistici. Tra questi, i più caratteristici sono:

1. $(P x \geq 0) \varphi(x)$ è sempre vero, perché ogni insieme ha misura almeno 0;
2. se $s < r$, allora $(P x \geq r) \varphi(x) \rightarrow (P x \geq s) \varphi(x)$ è valido, poiché dalla misura almeno r segue immediatamente che è anche almeno s ;
3. se vogliamo riflettere l'additività della misura, scriviamo uno schema che dice: se $\varphi(x)$ e $\psi(x)$ sono formule tali che il loro insieme di soluzioni è “quasi disgiunto” e quasi esaustivo, allora le misure si sommano; formalmente, un tipico assioma è:

$$(P x \geq r) \varphi(x) \wedge (P x \geq s) \psi(x) \wedge (P x \geq 1) (\neg \varphi(x) \vee \neg \psi(x)) \rightarrow (P x \geq \min(1, r + s)) (\varphi(x) \vee \psi(x)).$$

Questo dice che, se gli insiemi di soluzioni di φ e ψ sono praticamente disgiunti (la terza congiunzione afferma che quasi ovunque non è vero che entrambi valgono), allora la misura dell'unione è almeno la somma delle misure, troncata a 1.

4. un assioma di monotonicità del tipo: se $\forall x (\varphi(x) \rightarrow \psi(x))$ è vero, allora $(P x \geq r) \varphi(x) \rightarrow (P x \geq r) \psi(x)$, che riflette il fatto che l'insieme di soluzioni di φ è contenuto in quello di ψ , e quindi non può avere misura maggiore.

Il teorema di completezza che Keisler dimostra dice, in sostanza, che questa logica è completa rispetto a una classe di strutture misurate ben specifica, quella delle cosiddette misure di Loeb su iperfinito. Qui entra in gioco la teoria dei numeri non standard e delle ultrapotenze. L'idea è costruire una classe di domini che, pur essendo “infiniti” in senso standard, sono visti internamente come “iperfiniti” rispetto a un modello non standard dei numeri naturali. Su tali domini, possiamo definire una funzione di conteggio che, dal punto di vista interno, è finita, e da essa ricavare una misura interna finitamente additiva. La tecnica di Loeb mostra che questa misura interna, opportunamente “standardizzata”, genera una misura esterna σ -additiva sul dominio standard, chiamata misura di Loeb.

In dettaglio, prendiamo una famiglia di strutture finite $M_i = (D_i, I_i)$ per $i \in \mathbb{N}$, e consideriamo un ultrafiltro non principale U su $\mathbb{N}$. L'ultraprodotto $M^* = \prod_i M_i / U$ è un modello non standard che, dal punto di vista interno, ha un dominio D^* che è “iperfinito”: c'è un elemento $N^* \in \mathbb{N}$ (*l'estensione non standard dei naturali*) che è infinito rispetto a $\mathbb{N}$, e una biiezione interna tra D e $\{1, 2, \dots, N\}$. Esiste dunque una funzione di conteggio interna $\text{card_int}: P_int(D) \rightarrow \mathbb{N}$ che associa a ogni sottoinsieme interno $A \subseteq D$ la sua cardinalità interna $\text{card_int}(A)$; da questa si può definire una funzione di frequenza interna

$$\mu_int(A) = \text{card_int}(A) / \text{card_int}(D^*),$$

che a ogni sottoinsieme interno A assegna un “numero” in $[0, 1]$, *l'intervallo non standard dei reals esteso*. Questa μ_int è finitamente additiva ma solo internamente; la σ -addittività non ha nemmeno senso a priori, perché è definita solo su sottoinsiemi interni di D . Loeb dimostra che esiste una procedura di “standard part” che associa a ogni numero non standard $x \in {}^*[0, 1]$ un numero reale $\text{st}(x) \in [0, 1]$, e che se definiamo

$$\mu_L(A) = \text{st}(\mu_int(A))$$

per tutti i sottoinsiemi A di D^* che sono approssimazione di sottoinsiemi interni (si usa la saturazione e si definisce una σ -algebra di sottoinsiemi “misurabili” di D^*), otteniamo una misura σ -additiva completa μ_L su una σ -algebra Σ_L che contiene tutti i sottoinsiemi che ci interessano. Questa μ_L è la misura di Loeb associata alla misura interna μ_int .

Ora, una struttura di Loeb è una quadrupla $M_L = (D, I, \mu_L)$, dove D^* è il dominio iperfinito, I^* è l'interpretazione non standard del linguaggio L , e μ_L è la misura di Loeb sul dominio. In un tale modello, le espressioni $(P x \geq r) \phi(x)$ possono essere interpretate direttamente rispetto a μ_L . Keisler mostra che, dato ogni insieme di formule della logica con quantificatori probabilistici consistente rispetto al suo sistema assiomatico, esiste una struttura di Loeb in cui tutte queste formule sono soddisfatte. Questo è il teorema di completezza: la deducibilità coincide con la validità in tutte le strutture di Loeb.

Lo schema di dimostrazione segue il pattern di Henkin e della teoria dei modelli classica, arricchito dalla tecnologia non standard. Si parte da un insieme di formule Σ del linguaggio con i quantificatori $(P x \geq r)$ che sia consistente. Per estendere Σ a una teoria completa, si usano tecniche di completamento alla Henkin, aggiungendo costanti di Henkin e assicurando che ogni formula esistenziale venga testimoniata. Si ottiene una teoria massimale T che è ancora consistente. A questo punto, si vuole costruire una struttura M_T che la soddisfi. Invece di farlo direttamente, Keisler costruisce una catena diretta di strutture finite, ciascuna delle quali approssima un frammento di T , e poi forma un ultraproducto M^* di questa catena. Nel processo, si codifica in modo “quasi uniforme” la frequenza delle soluzioni dei predicati e delle formule con variabile libera, in modo che, nel limite ultraproiettivo, le misure interne soddisfino i vincoli espressi dai quantificatori probabilistici. La difficoltà è garantire la coerenza di tutte queste approssimazioni simultaneamente, ma qui interviene la compattezza della logica del primo ordine e la proprietà degli ultrafiltri non principali: ogni vincolo finito è soddisfatto da qualche stadio, e l'ultrafiltrato “satura” tutti i vincoli compatibili.

Una volta ottenuto M^* con misura interna μ_{int} , si passa a μ_L mediante la costruzione di Loeb. Infine, si verifica che, relativamente a μ_L , tutte le formule con quantificatori $(P x \geq r)$ di T sono soddisfatte, grazie al fatto che i valori interni delle frequenze erano calibrati per codificare esattamente i valori desiderati. La delicatezza tecnica sta nel mostrare che gli assiomi di Keisler sono sufficienti per garantire che queste calibrazioni siano compatibili con la σ -additività che nasce dopo la standardizzazione. L'uso della non standard analysis consente di evitare complicazioni dirette con successioni di insiemi e limiti di misure; tutto viene incapsulato nella struttura iperfinita, dove la finitezza interna permette di lavorare con semplici conteggi e rapporti interi.

Se spostiamo ora l'attenzione dalle misure sul dominio alle misure su insiemi di strutture, entriamo nel territorio del lavoro di Hoover. Qui il punto di partenza è un insieme di strutture per un linguaggio L , e una misura di probabilità μ definita su di esso. Una misura μ su strutture può essere interpretata come un “pattern” di incertezza sulla verità di enunciati del linguaggio: la probabilità che una formula $\phi(x_1, \dots, x_n)$ sia vera di una certa tupla di elementi dipende dalla distribuzione di μ sulle strutture. L'idea è chiedersi in che senso questa μ sia “definibile” in termini del linguaggio e quali siano le condizioni sotto cui essa si comporta in modo regolare, ad esempio essendo invariante sotto permutazioni del dominio, cioè scambiabile.

Una misura μ su strutture di L è detta scambiabile se, intuitivamente, non distingue tra gli elementi del dominio se non per le proprietà che il linguaggio è in grado di esprimere. Formalmente, prendiamo un insieme di strutture $M = (D, I)$ con D numerabile, e consideriamo, per ciascuna permutazione finita π di D , l'applicazione che manda M in M^π , ottenuta permutando il dominio e trascinando l'interpretazione. La misura μ è scambiabile se μ è invariante rispetto a queste permutazioni: per ogni insieme misurabile A di strutture, $\mu(A) = \mu(\{ M^\pi \mid M \in A \})$ per ogni π . In altre formulazioni, si costruisce un oggetto più astratto: una probabilità su configurazioni infinite di valori di predicati su tuple indicizzate da $\mathbb{N}$, e si richiede che questa probabilità sia invariante rispetto a tutte le permutazioni di $\mathbb{N}$.

Il risultato centrale di Hoover, in una forma semplificata, afferma che misure scambiabili su strutture, sotto ipotesi aggiuntive di definibilità, ammettono una rappresentazione come miscele di misure prodotto. L'analogia con il teorema di de Finetti per sequenze di variabili Bernoulli è diretta: là, una misura scambiabile su $\{0,1\}^{\mathbb{N}}$ può essere rappresentata come integrale di misure prodotto i.i.d. parametro θ , con θ distribuito secondo una misura a priori; qui, una misura scambiabile su strutture di L può essere vista come integrale di misure prodotto su configurazioni di fatti atomici, con rispetto a un parametro “globale” che codifica la frequenza delle relazioni.

Per capire il contenuto, consideriamo un linguaggio L molto semplice, con un solo predicato monadico R . Una struttura M di questo tipo, con dominio D numerabile, è determinata da una funzione $f_M : D \rightarrow \{0,1\}$ che indica quali elementi soddisfano R . Una misura μ su queste strutture induce una misura $\tilde{\mu}$ su $\{0,1\}^{\mathbb{N}}$, identificando D con $\mathbb{N}$. Scambiabilità qui significa precisamente che $\tilde{\mu}$ è invariante sotto permutazioni finite di $\mathbb{N}$: $\tilde{\mu}(A) = \tilde{\mu}(\pi(A))$ per ogni insieme misurabile A e permutazione π . Il teorema di de Finetti ci dice che esiste una misura π su $[0,1]$ tale che per ogni insieme misurabile $B \subseteq \{0,1\}^{\mathbb{N}}$,

$$\tilde{\mu}(B) = \int 0^{\sum x_n} 1^{\sum (1-x_n)} \mu(\theta) d\pi(\theta),$$

dove μ_θ è la misura prodotto i.i.d. su $\{0,1\}^{\mathbb{N}}$ con parametro θ : per ogni sequenza $(x_n)_n$,

$$\mu_\theta(\{(x_n)_n\}) = \prod_n \theta^{x_n} (1-\theta)^{1-x_n}.$$

Questa è una rappresentazione come miscela di misure prodotto. Hoover generalizza questo quadro a linguaggi con relazioni di arità superiore, ottenendo miscele di misure prodotto su spazi più complicati (per esempio, spazi di matrici binarie per relazioni binarie, spazi di funzioni su tuple per relazioni n -arie, ecc.). Il parametro θ non è più un numero reale singolo, ma un oggetto più strutturato, qualcosa come una famiglia di parametri (θ_R) , uno per ciascun predicato e per ciascun pattern di indici.

Il teorema di Hoover può essere enunciato, in forma concettuale, così: sia μ una misura scambiabile su strutture di un linguaggio L , con dominio numerabile. Allora esiste uno spazio parametrico Θ , una misura a priori ν su Θ , e per ogni $\theta \in \Theta$ una misura prodotto μ_θ su strutture di L (in cui i fatti atomici sono indipendenti, dato θ), tali che

$$\mu = \int \mu_\theta d\nu(\theta).$$

La “prodotto” qui è da intendersi in un senso tecnico: condizionatamente a θ , i fatti atomici come $R(a_1, \dots, a_n)$ diventano indipendenti, con probabilità determinate da θ . Si tratta di una generalizzazione delle rappresentazioni alla Aldous-Hoover per array scambiabili, che a loro volta estendono de Finetti alle famiglie indicizzate da coppie, triple, ecc., e catturano strutture casuali come grafi, ipergrafi, relazioni, ecc. Il contenuto logico di questo quadro è che misure scambiabili su strutture logiche possono essere sempre viste come miscele di “modelli indipendenti”, dove “indipendente” significa che, una volta fissato il parametro globale θ , i fatti atomici sono probabilisticamente indipendenti.

La nozione di definibilità entra nel quadro per limitare la complessità di Θ e la forma delle μ_θ . Una misura μ su strutture di L è definibile se, per ogni formula $\phi(x)$ e ogni $\varepsilon > 0$, esiste una formula priva di quantificatori $\psi(x)$ tale che la probabilità che $\phi(x)$ differisca da $\psi(x)$ in misura più di ε è piccola. Più precisamente, c'è una nozione di “quasi ovunque” che viene formalizzata tramite μ : la condizione è qualcosa come

$$\mu(\{M \mid M \models \forall x (\varphi(x) \leftrightarrow \psi(x))\}) \geq 1 - \varepsilon.$$

Questa proprietà dice che, dal punto di vista della misura μ , ogni formula può essere approssimata arbitrariamente bene da una combinazione booleana di fatti atomici. In altre parole, la struttura logica addizionale portata dai quantificatori non introduce complessità probabilistica che non possa essere resa con formule quantificatore-free. Hoover dimostra che, sotto scambiabilità e definibilità, la rappresentazione come miscela di misure prodotto assume una forma particolarmente regolare, che si collega a quanto visto con Scott-Krauss e con la rappresentazione delle misure logiche.

Nel quadro che stiamo costruendo, possiamo allora vedere una linea di continuità: Gaifman fornisce misure su spazi di modelli che rispettano la struttura logica; Carnap colloca la conferma induttiva in un modello parametrico binomiale e ne mostra la rappresentabilità bayesiana; Scott e Krauss rappresentano misure su formule come misure su spazi di probabilità; Keisler arricchisce la logica con quantificatori probabilistici e ottiene completezza rispetto a strutture di Loeb; Hoover e i risultati su scambiabilità mostrano che misure “simmetriche” su strutture possono essere viste come miscele di mondi indipendenti. Tutto converge verso l’idea che la probabilità, quando è applicata alla logica, non è un semplice grado di credenza epistemico, ma una struttura matematica che vive nello stesso mondo degli ultrafiltri, dei modelli, delle ultrapotenze, delle algebre di misura.

Se ora, su questa base fondazionale, vogliamo vedere come la probabilità entri nel livello proposizionale e modale in modalità più operative e computazionali, il passo naturale è tornare alla logica di Fagin-Halpern-Megiddo e a Nilsson, ma questa volta con una lente più tecnica. Nel quadro di Fagin-Halpern-Megiddo, l’elemento originale è l’introduzione di termini probabilistici nel linguaggio stesso. Si consideri un insieme finito di proposizioni atomiche $p_1, \dots, p_n$, e un insieme W di mondi possibili. Una valutazione classica π associa a ogni mondo $w \in W$ e ogni proposizione p_i un valore di verità $\pi(w, p_i) \in \{0,1\}$. Una funzione $\mu: W \rightarrow [0,1]$ definisce una distribuzione di probabilità sui mondi (normalizzata: $\sum_{w \in W} \mu(w) = 1$). La probabilità di una formula proposizionale φ , interpretata nel modello (W, π, μ) , è definita come

$$\Pr(\varphi) = \sum_{w \in W, (W, \pi, w) \models \varphi} \mu(w),$$

cioè la somma delle probabilità dei mondi in cui φ è vera. La logica estende il linguaggio proposizionale con espressioni come $\Pr(\varphi) \geq \alpha$, dove α è un numero razionale in $[0,1]$, e con combinazioni lineari di termini probabilistici. Nella versione completa, le formule possono contenere vincoli lineari del tipo

$$a_1 \Pr(\varphi_1) + \dots + a_k \Pr(\varphi_k) \geq b,$$

dove a_i, b sono razionali. La semantica dice che un modello (W, π, μ) soddisfa tale formula se il vincolo lineare sui numeri reali risulta vero quando sostituiamo a ciascuna $\Pr(\varphi_i)$ il valore definito sopra.

La domanda computazionale fondamentale è: data una formula Φ in questo linguaggio, esiste un modello (W, π, μ) che la soddisfa? Fagin, Halpern e Megiddo mostrano che questo problema è decidibile e ne caratterizzano la complessità. Il trucco tecnico è ridurre il problema logico a un problema di fattibilità di un sistema di disequazioni lineari. Osserviamo che ogni formula φ_i è proposizionale su n atomi, quindi corrisponde a un sottoinsieme di 2^n assegnazioni di verità. Ogni mondo w può essere identificato con una di queste assegnazioni, e quindi possiamo considerare direttamente la probabilità di ciascuna assegnazione come una variabile $x_j, j = 1, \dots, 2^n$, soggetta ai vincoli:

1. $x_j \geq 0$ per tutti j ;
2. $\sum_{j=1}^{2^n} x_j = 1$.

La probabilità $\Pr(\varphi_i)$ è allora la somma di tutte le x_j tali che l'assegnazione j rende vera φ_i . Quindi i vincoli $a_1 \Pr(\varphi_1) + \dots + a_k \Pr(\varphi_k) \geq b$ diventano vincoli lineari sulle x_j : basta scrivere ciascun $\Pr(\varphi_i)$ come $\sum_{j \in S_i} x_j$, dove S_i è l'insieme delle assegnazioni che soddisfano φ_i , e sostituire. La formula globale Φ può essere tradotta in un sistema $S(\Phi)$ di disequazioni lineari in 2^n variabili. La formula è soddisfacibile se e solo se il sistema $S(\Phi)$ ha una soluzione reale non negativa con somma 1. Esistono algoritmi di programmazione lineare che decidono la fattibilità di un tale sistema in tempo polinomiale nel numero di variabili e vincoli, ma nel nostro caso il numero di variabili è esponenziale nel numero di atomi n . Ne segue che la soddisfacibilità è, in generale, di complessità elevata (tipicamente EXPTIME). Tuttavia, la possibilità stessa di ridurre la logica a programmazione lineare è concettualmente importante: la logica probabilistica proposizionale si rivela isomorfa a un poliedro convesso nello spazio delle probabilità sulle assegnazioni atomiche.

Nilsson, in un contesto affine ma con un'angolatura diversa, parte da un insieme di formule proposizionali $\varphi_1, \dots, \varphi_m$, ciascuna accompagnata da un intervallo di probabilità $[l_i, u_i]$, e pone il problema: esiste una distribuzione di probabilità p sulle assegnazioni $v_1, \dots, v_{2^n}$ tale che per ogni i ,

$$l_i \leq \Pr(\varphi_i) \leq u_i?$$

Qui $\Pr(\varphi_i)$ è, come prima, la somma delle probabilità $p(v_j)$ per j tali che $v_j \models \varphi_i$. L'insieme delle distribuzioni p compatibili con i vincoli forma un politopo convesso K in $\mathbb{R}^{2^n}$, definito da:

1. $p_j \geq 0$ per tutti j ;
2. $\sum_j p_j = 1$;
3. per ogni i , $l_i \leq \sum_{j \in S_i} p_j \leq u_i$.

La struttura di K è quella di un politopo definito da disequazioni lineari. L'inferenza probabilistica in questo quadro consiste nel determinare, per una nuova formula ψ , l'intervallo più stretto $[L, U]$ tale che, per ogni distribuzione p compatibile con i vincoli, si abbia $L \leq \Pr(\psi) \leq U$. Questo equivale a risolvere due problemi di programmazione lineare: minimizzare $\sum_{j \in S_\psi} p_j$ su K e massimizzarla su K . La complessità è in generale esponenziale in n , ma per problemi con pochi atomi o con struttura speciale si possono avere algoritmi efficienti.

Si può vedere subito il parallelismo con la rappresentazione di Scott-Krauss: l'insieme delle misure logiche (su formule proposizionali) che rispettano i vincoli di Nilsson è anch'esso un politopo convesso in uno spazio di dimensione $2^n - 1$. Il trattamento misura-teoretico astratto di Scott-Krauss si specializza nel caso proposizionale a una geometria convessa concreta; Fagin-Halpern-Megiddo e Nilsson sono due modi diversi di muoversi all'interno di quella geometria: nel primo si formulano vincoli lineari tra probabilità di formule, nel secondo si parte direttamente da intervalli assegnati.

Una volta accettato questo radicamento geometrico e misura-teoretico, la strada verso le logiche modali e temporali probabilistiche è naturale: al posto di uno spazio discreto di mondi, consideriamo sistemi dinamici come catene di Markov, e al posto di assegnazioni statiche di probabilità alle formule, consideriamo probabilità che certe proprietà temporali siano soddisfatte lungo i cammini. Qui entra ad esempio PCTL (Probabilistic Computational Tree Logic), in cui formule come $P_{\geq r}[\varphi \cup \psi]$ esprimono che la probabilità che, lungo un cammino generato dalla catena di Markov, ψ diventi vera prima o poi e φ rimanga vera fino a quel momento è almeno r . La semantica assegna a ogni stato s di una catena di Markov $M = (S, P, L)$ la probabilità $\Pr_s$ che un certo insieme di cammini

soddisfi una formula di cammino; questa Pr_s è definita come misura sullo spazio dei cammini, dove la misura è costruita a partire dalla matrice di transizione P mediante un prodotto infinito di probabilità condizionali: se un cammino è una sequenza $s_0, s_1, s_2, \dots$, la probabilità di questa sequenza (per una catena senza stati assorbenti, o comunque con definizione adeguata) è

$$Pr(\{s_0, s_1, s_2, \dots\}) = \prod_{i=0}^{\infty} P(s_i, s_{i+1}).$$

Lo spazio dei cammini è un prodotto countabile $S^{\mathbb{N}}$, e su di esso si costruisce la misura cilindrica generata dalle probabilità finite: l'insieme dei cammini che iniziano per $s_0, \dots, s_n$ ha misura $P(s_0, s_1) \cdots P(s_{n-1}, s_n)$. La probabilità Pr_s di un evento (insieme di cammini) A è la misura $\mu_s(A)$ associata alla misura cilindrica con radice in s . Il verificarsi di $P\{\geq r\}[\varphi \cup \psi]$ in s significa che $\mu_s(A_{\{\varphi \cup \psi\}}) \geq r$, dove $A_{\{\varphi \cup \psi\}}$ è l'insieme dei cammini che soddisfano la proprietà temporale $\varphi \cup \psi$.

Risultati come la preservazione della soddisfazione di formule PCTL sotto bisimulazione probabilistica mostrano che tali logiche hanno un'invarianza robusta rispetto alla struttura probabilistica: due stati sono probabilisticamente bisimili se non solo hanno le stesse etichette atomiche, ma anche lo stesso “profumo probabilistico” verso le classi di equivalenza. In questo senso, la logica probabilistica modale e temporale porta nel primo piano un concetto di equivalenza di stati che è profondamente misura-teoretico: non basta che due stati abbiano lo stesso insieme di proprietà atomiche; devono avere le stesse distribuzioni di probabilità sui loro futuri.

Tutto questo ci riporta, per un altro percorso, al quadro generale: la probabilità in logica si manifesta come misura su insiemi di modelli (Gaifman), come misura logica su algebre di Lindenbaum-Tarski (Scott-Krauss), come distribuzione su parametri per famiglie induttive (Carnap), come misura su domini con quantificatori (Keisler e Loeb), come misura su strutture scambiabili (Hoover), come politopi convessi per limiti proposizionali (Nilsson, Fagin-Halpern-Megiddo), come misure su spazi di cammini in sistemi dinamici (PCTL). Ogni volta, gli oggetti tecnici sono diversi – spazi di Stone, ultrapotenze, misure di Loeb, politopi, catene di Markov – ma l'architettura resta sorprendentemente unitaria: una algebra di eventi logici, una misura che la popola, e vincoli logici che plasmano la geometria di quella misura.

Per chiarire con rigore la misura di Loeb, partiamo da zero nel linguaggio della non standard analysis. Si considera un'estensione non standard ${}^*\mathbb{N}$ dell'insieme dei numeri naturali $\mathbb{N}$, ottenuta via ultraprodotto: si prende il prodotto $\prod_{i \in I} \mathbb{N}$ e lo si quozienta rispetto a un ultrafiltro non principale U , ottenendo un campo di numeri “iperinteri” in cui $\mathbb{N}$ si immerge diagonalmente ma compaiono anche elementi infiniti, cioè maggiori di ogni $n \in \mathbb{N}$. Analogamente, si costruisce ${}^*\mathbb{R}$ come ultraprodotto di copie di $\mathbb{R}$, ottenendo un campo ordinato non archimedeo che contiene $\mathbb{R}$ come sottoinsieme di elementi “standard”, e un anello di “infinitesimi” costituito da elementi ε tali che $|\varepsilon| < 1/n$ per ogni n standard. In questo contesto, un insieme interno di un certo universo non standard è, grosso modo, un oggetto costruito come immagine nell'ultraprodotto di una famiglia di insiemi standard; in particolare, un sottoinsieme interno A di uno spazio X è uno di questi oggetti generati da processi di ultraprodotto.

Applicata alle strutture di modelli, la tecnica funziona così. Si consideri una sequenza di strutture finite $M_i = (D_i, I_i)$, $i \in \mathbb{N}$, ciascuna con un dominio finito D_i . Si forma l'ultraprodotto $M^* = \prod_i M_i / U$. Il dominio D^* di M^* è l'ultraprodotto dei D_i , ed è un insieme interno che, dal punto di vista non standard, è “iperfinito”: cioè esiste un iperintero $N^* \in \mathbb{N}$ e una biiezione interna $f: D \rightarrow \{1, \dots, N\}$. Questo permette di definire una funzione di conteggio interna: per ogni sottoinsieme interno $A \subseteq D$, definito anch'esso come ultraprodotto di sottoinsiemi $A_i \subseteq D_i$, poniamo

$\text{card_int}(A) = |A|_{\text{int}} = \text{numero interno di elementi di } A,$

cioè la cardinalità nella struttura non standard. Poiché D^* ha cardinalità interna N^* , è naturale definire una funzione di frequenza interna:

$$\mu_{\text{int}}(A) = \text{card_int}(A) / \text{card_int}(D^*) \in {}^*[0,1].$$

Questa μ_{int} è finitamente additiva sugli insiemi interni: se A e B sono interni e disgiunti, $\mu_{\text{int}}(A \cup B) = \mu_{\text{int}}(A) + \mu_{\text{int}}(B)$, perché il conteggio interno è additivo. Inoltre, $\mu_{\text{int}}(D^*) = 1$ per definizione. Dal punto di vista interno, μ_{int} è quindi una misura di probabilità su un'anello (anzi un'algebra) di insiemi interni. Il problema è che μ_{int} prende valori in $[0,1]$, *non in $[0,1]$, e non è definita, a priori, su tutti i sottoinsiemi esterni di D* , né abbiamo una nozione di σ -additività esterna.

Il passo centrale di Loeb è definire la “parte standard” di μ_{int} . Per ogni $x \in {}^*\mathbb{R}$ finito (cioè non infinito), esiste un unico numero reale $\text{st}(x) \in \mathbb{R}$ tale che $x - \text{st}(x)$ è infinitesimo. La funzione $\text{st}: {}^*\mathbb{R}\{\text{finito}\} \rightarrow \mathbb{R}$ è un omomorfismo parziale che “collassa” gli infinitesimi a 0. Se A è un insieme interno di D^* , $\mu_{\text{int}}(A)$ è un numero interno in ${}^*[0,1]$, quindi finito, e possiamo considerare $\text{st}(\mu_{\text{int}}(A)) \in [0,1]$. L'idea è definire una nuova funzione μ_L su una σ -algebra più grande che contenga gli insiemi interni, tale che $\mu_L(A)$ coincida con $\text{st}(\mu_{\text{int}}(A))$ sugli insiemi interni, e sia estesa in modo σ -additivo.

Formalmente, si definisce una pre-misura su una algebra A_0 di sottoinsiemi di D^* , prendendo come algebra quella generata dagli insiemi interni. Su A_0 , si pone

$$\mu_0(B) = \inf \{ \sum_i \text{st}(\mu_{\text{int}}(A_i)) : B \subseteq \bigcup_i A_i, A_i \text{ interni} \}.$$

Cioè $\mu_0(B)$ è l'infimo di tutte le somme delle parti standard di $\mu_{\text{int}}(A_i)$ su famiglie di insiemi interni che ricoprono B . Questo è un procedimento nello stile di Carathéodory: definire un'outer measure a partire da una misura preesistente su un sottoinsieme di insiemi. Si dimostra che μ_0 è una outer measure su $P(D)$, e quindi esiste una σ -algebra Σ_L di sottoinsiemi di D che sono misurabili alla Carathéodory, rispetto alla quale μ_0 è σ -additiva. La misura di Loeb μ_L è poi la restrizione di μ_0 a Σ_L . Loeb dimostra che, per ogni insieme interno A , A è misurabile e $\mu_L(A) = \text{st}(\mu_{\text{int}}(A))$. Questo garantisce che non abbiamo perso l'informazione originale: la misura interna si riflette esattamente nella misura di Loeb, ma quest'ultima è ora una misura classica (σ -additiva, completa, definita su una σ -algebra esterna).

Applicata al quadro di Keisler, questa costruzione significa che le frequenze interne di verità di una formula $\varphi(x)$ in M^* diventano, tramite st , la misura reale $\mu_L(\{a \in D^* \mid M^* \models \varphi(a)\})$. Quando scriviamo $(P \ x \geq r) \varphi(x)$, lo interpretiamo come $\mu_L(\{a \mid \varphi(a)\}) \geq r$. Gli assiomi di Keisler sono pensati per essere veri in tutte le strutture di Loeb: l'additività, la monotonia, l'ovvietà di $(P \ x \geq 0) \varphi(x)$, ecc. Il teorema di completezza afferma che se una formula con quantificatori probabilistici è vera in tutte le strutture di Loeb, allora è dimostrabile dal sistema assiomatico; reciprocamente, ogni formula dimostrabile è valida in tutte le strutture di Loeb. Hoover affina questa analisi dimostrando un teorema di completezza analitica per logiche con quantificatori probabilistici, in cui non ci si appoggia più esclusivamente a strutture non standard, ma si ottengono modelli “ben comportati” come sottoinsiemi analitici di $\mathbb{R}$ con relazioni e funzioni Borel misurabili; l'idea è che la rappresentazione di misure e quantificatori probabilistici può essere realizzata all'interno della teoria classica della misura sui reali, con strutture di modelli che vivono dentro $\mathbb{R}$ e hanno tutte le buone proprietà analitiche desiderate Cambridge University Press & Assessment. In questo modo, la connessione con la teoria classica della misura (Lebesgue, Borel, prodotto di misure, ecc.) è ancora più stretta: i modelli non sono entità astratte, ma insiemi analitici nello spazio reale.

Passando alle strutture scambiabili, l'asse fondamentale si sposta dalla misura sul dominio alla misura sullo spazio delle strutture stesse. L'analogo diretto del teorema di de Finetti per sequenze di variabili aleatorie su $\{0,1\}$ è il teorema di Aldous–Hoover per matrici (o array) di variabili casuali indicizzate da coppie (o da tuple più generali). Consideriamo una matrice infinita di variabili casuali $(X_{ij})_{i,j \in \mathbb{N}}$ a valori in qualche spazio misurabile, per esempio $\{0,1\}$ o $\mathbb{R}$. Si dice che la distribuzione congiunta di (X_{ij}) è invariante per permutazioni finite di righe e colonne se, per ogni permutazione finita π di $\mathbb{N}$, la legge di (X_{ij}) è la stessa di quella di $(X_{\pi(i)\pi(j)})$. Il teorema di Aldous–Hoover afferma che ogni distribuzione con questa simmetria può essere rappresentata come segue: esistono variabili casuali indipendenti

$U \in [0,1]$, $U_i \in [0,1]$ per $i \in \mathbb{N}$, $V_j \in [0,1]$ per $j \in \mathbb{N}$, $W_{ij} \in [0,1]$ per $i,j \in \mathbb{N}$,

tutte i.i.d. uniformi su $[0,1]$, e una funzione misurabile $f: [0,1]^4 \rightarrow S$ (dove S è lo spazio dei valori di X_{ij}) tale che

$$X_{ij} = f(U, U_i, V_j, W_{ij})$$

in legge. Cioè, a meno di uguaglianza in distribuzione, ogni matrice scambiabile può essere generata partendo da una variabile globale U comune, da variabili di riga U_i , di colonna V_j e da variabili specifiche d'entrata W_{ij} , tutte indipendenti e uniformi, attraverso una funzione misurabile f . Dal punto di vista della logica, questo dice che una struttura scambiabile con una relazione binaria (per esempio, un grafo aleatorio) può essere vista come una funzione di rumore indipendente e di alcuni parametri globali, di riga e di colonna, che codificano le “caratteristiche latenti” dei nodi e dell'intero grafo.

Nel quadro delle strutture logiche, Aldous–Hoover viene usato per rappresentare misure scambiabili su modelli infiniti. Se abbiamo un linguaggio con un solo simbolo binario R (pensiamo a grafi non orientati o a grafi orientati), una misura scambiabile su strutture di L con dominio numerabile equivale alla legge di un array di variabili casuali $X_{ij} = 1_{\{R(i,j)\}}$. La rappresentazione di Aldous–Hoover ci dice che questa legge è generata da un oggetto latente U globale, da etichette latenti U_i per i nodi, e da un rumore W_{ij} per le coppie, attraverso una funzione f . Per grafi non orientati senza loop, si aggiungono simmetrie: $X_{ij} = X_{ji}$, e si richiede che f sia simmetrica in U_i, V_j . Dal punto di vista probabilistico, la struttura scambiabile è quindi una miscela (integrale rispetto alla legge di U e degli U_i) di modelli indipendenti condizionati a questi parametri latenti. Questo è precisamente il contenuto concettuale che Hoover porta nel contesto logico: misure scambiabili su modelli sono miscele di prodotti condizionati a parametri; nel linguaggio della teoria dei modelli, si tratta di misure su spazio di strutture che, una volta disintegrate rispetto a determinati invarianti, hanno struttura di prodotto sugli eventi atomici.

La definibilità, in questo contesto, assume una forma raffinata: una misura μ su strutture di L è definibile se, per ogni formula $\varphi(x_1, \dots, x_n)$ e ogni $\varepsilon > 0$, esiste una formula priva di quantificatori $\psi(x_1, \dots, x_n)$ tale che, per “quasi tutte” le strutture M rispetto a μ , gli insiemi di tuple che soddisfano φ e ψ coincidono a meno di un insieme di misura piccola. In simboli, qualcosa come:

$$\mu(\{M \mid \mu_M(\{a \in D^n \mid M \models \varphi(a) \leftrightarrow \psi(a)\}) \geq 1 - \varepsilon\}) \geq 1 - \varepsilon,$$

dove μ_M è la misura sul dominio di M (se ne abbiamo una) o una misura di conteggio normalizzata sulle n -uple. Questo tipo di condizione garantisce che la complessità portata dai quantificatori possa essere “assorbita” in strutture più elementari: la misura μ “vede” il mondo quasi interamente attraverso fatti atomici e combinazioni finitarie di essi. Hoover, combinando scambiabilità e definibilità, mostra che la rappresentazione come miscela di prodotti assume una forma estremamente

regolare e che molte strutture logiche complesse possono essere comprese come “sistemi random i.i.d. sopra parametri latenti”, almeno dal punto di vista della misura.

Questo quadro si collega in modo naturale alle fondazioni categoriali moderne della probabilità. Una categoria di Markov è, grosso modo, una categoria monoidale simmetrica in cui gli oggetti rappresentano tipi di sistemi e i $X \rightarrow Y$ rappresentano “canali probabilistici” o “trasformazioni stocastiche” tra questi sistemi. Un esempio concreto è la categoria *FinStoch*, dove gli oggetti sono insiemi finiti e un morfismo $f: X \rightarrow Y$ è una matrice stocastica, cioè una famiglia di distribuzioni di probabilità su Y indicizzate da elementi di X : formalmente una funzione $f: X \times Y \rightarrow [0,1]$ tale che, per ogni x , $\sum_y f(x,y) = 1$. Un altro esempio, più generale, è la categoria *Meas*, dove gli oggetti sono spazi misurabili e i morfismi $X \rightarrow Y$ sono kernel di Markov, cioè funzioni $K: X \times \Sigma_Y \rightarrow [0,1]$ tale che, per ogni x , $K(x, \cdot)$ è una misura di probabilità su Y , e per ogni insieme misurabile $B \subseteq Y$, la funzione $x \mapsto K(x,B)$ è misurabile.

La monade di Girard formalizza proprio questo concetto di “trasformazione probabilistica”: per ogni spazio misurabile X , Girard associa l’insieme $G(X)$ delle misure di probabilità su X , dotato di una σ -algebra naturale (la minore σ -algebra che rende misurabile il mapping $v \mapsto v(B)$ per ogni B misurabile di X). Un morfismo misurabile $f: X \rightarrow Y$ induce un morfismo $G(f): G(X) \rightarrow G(Y)$ spedendo una misura v su X nella sua immagine diretta $f_*(v)$ su Y . La monade di Girard (G, η, μ) combina questo operatore con una unità $\eta_X: X \rightarrow G(X)$, che manda ogni x nella delta di Dirac δ_x , e con una moltiplicazione $\mu_X: G(G(X)) \rightarrow G(X)$ che integra misure di misure in misure semplici: data una misura Π su $G(X)$, $\mu_X(\Pi)$ è la misura su X definita da $\mu_X(\Pi)(A) = \int v(A) d\Pi(v)$. Questo quadro astratto incapsula precisamente l’idea di “scelta randomizzata” e di “combinazione di misure” che abbiamo usato ovunque nel discorso su logiche probabilistiche: la misura di Gaifman su modelli, la miscela di de Finetti e Aldous–Hoover, la disintegrazione in parametri latenti di Hoover, la media bayesiana di Carnap–Stegmüller sono tutte istanze della stessa operazione μ_X : trasformare misure su misure in misure semplici.

Nella prospettiva delle categorie di Markov, la logica probabilistica può essere vista come la teoria delle formule invarianti rispetto a certe classi di morfismi probabilistici. Ad esempio, una formula del primo ordine con quantificatori probabilistici può essere interpretata come una certa proprietà di un morfismo $X \rightarrow 2$ (dove 2 è lo spazio booleano $\{0,1\}$) che rimane stabile sotto composizioni con altri morfismi e prodotti. In un lavoro recente, risultati alla Aldous–Hoover vengono persino riformulati completamente nel linguaggio delle categorie di Markov: una famiglia di variabili casuali scambiabili diventa un morfismo da un oggetto “indice” (una sorta di oggetto che rappresenta $\mathbb{N}$ o $\mathbb{N} \times \mathbb{N}$) a un oggetto target, e l’invarianza sotto permutazioni si traduce in condizioni di simmetria categorica; la rappresentazione come funzione di variabili latenti indipendenti diventa un fattore di questo morfismo attraverso un prodotto di copie dell’oggetto unità con sé stesso arxiv.org/abs/1808.07441. Questa riformulazione mostra che la struttura scambiabile non è un incidente dei modelli insiemistici, ma è qualcosa che può essere espresso in termini puramente categoriali, come proprietà di un morfismo in una categoria probabilistica astratta.

A questo punto il ponte con le logiche di ordine superiore e con i linguaggi probabilistici (probabilistic programming languages) è naturale. Un λ -calcolo tipato probabilistico può essere visto come una teoria interna in una categoria con monade di Girard: i tipi sono oggetti, i termini sono morfismi, il tipo “distribuzione su A ” è l’oggetto $G(A)$, e l’operatore di campionamento è la struttura di Kleisli della monade. Un termine $M: 1 \rightarrow G(A)$ rappresenta una distribuzione su A ; un termine $N: A \rightarrow G(B)$ rappresenta un kernel di Markov da A a B ; la composizione Kleisli di M e N rappresenta il comporre scelte randomizzate. Le equazioni del calcolo probabilistico (per esempio la legge di Fubini/tonelli al livello dei tipi) esprimono esattamente l’associatività della moltiplicazione della monade e la commutatività relativa al prodotto. Quando si dimostra che l’uguaglianza osservazionale di termini

in un λ -calcolo probabilistico non è decidibile, spesso lo si fa riducendo il problema al comportamento di macchine di Turing probabilistiche: si costruisce un termine che simula il processo randomizzato e si mostra che decidere se due distribuzioni coincidono per tutti gli input è indecidibile. È la stessa idea che compare nel teorema di Statman sull'ind decidibilità del λ -calcolo semplicemente tipato, portata nel dominio probabilistico.

La presenza di più costruzioni misura-teoretiche nella logica probabilistica — la misura sui modelli di Gaifman, le funzioni di conferma di Carnap, la rappresentazione di Scott e Krauss, i quantificatori probabilistici di Keisler, le misure scambiabili di Hoover — può dare l'impressione di una certa ridondanza. In realtà, queste teorie non sono duplicazioni, ma manifestazioni diverse di un'unica struttura profonda: la possibilità di interpretare la logica come teoria della misura e la misura come semantica logica. Ogni approccio illumina un aspetto diverso di questa identità, e ciò che sembra ripetersi è in realtà la ricorrenza di un principio strutturale che attraversa tutta la disciplina.

Il punto di partenza è che ogni linguaggio logico L induce naturalmente uno spazio di modelli Q , e ogni formula χ definisce un sottoinsieme $[\chi] \subseteq Q$. La domanda fondamentale è se esista una misura μ su Q tale che $\mu([\chi])$ rifletta la struttura logica di χ . Gaifman risponde affermativamente costruendo una misura su Q che rispetta i connettivi e i quantificatori, Carnap mostra che le funzioni di conferma sono casi particolari di misure su spazi parametrici, Scott e Krauss dimostrano che ogni misura logica su $LT(L)$ è rappresentabile come misura classica su uno spazio di probabilità, Keisler estende la logica con quantificatori probabilistici interpretati tramite misure di Loeb, Hoover mostra che le misure scambiabili su strutture sono rappresentabili come miscele di misure prodotto. La ripetizione non è un difetto: è la manifestazione di un teorema non scritto, secondo cui la logica probabilistica è una teoria della misura in più travestimenti.

La costruzione di Gaifman è la più diretta: si dota Q di una σ -algebra generata dagli insiemi $[\chi]$ e si costruisce una misura μ che rispetta la combinatoria logica. La parte apparentemente ridondante è che Scott e Krauss ripetono la stessa costruzione, ma a livello dell'algebra di Lindenbaum–Tarski $LT(L)$. In realtà, le due costruzioni sono duali: Gaifman lavora nello spazio dei modelli, Scott e Krauss nello spazio delle teorie complete. Il fatto che entrambe funzionino non è una coincidenza, ma un riflesso della dualità di Stone tra algebre booleane e spazi di ultrafiltri. La misura di Gaifman su Q induce una misura su $S(L)$, lo spazio delle teorie complete, e la misura di Scott–Krauss su $LT(L)$ induce una misura su uno spazio di modelli astratto. Le due costruzioni sono equivalenti a meno di isomorfismo di measure algebras. La duplicazione è solo apparente: è la stessa struttura vista da due lati della dualità.

La costruzione di Carnap sembra muoversi in un'altra direzione, ma in realtà è un caso particolare della stessa idea. La funzione di conferma $c_{\Lambda}(h|e)$ è una probabilità condizionata rispetto a una misura a priori τ su un parametro $\theta \in [0,1]$. La formula $c_{\Lambda}(h|e) = \int f(h,\theta) \Pr(e|\theta) d\tau(\theta) / \int \Pr(e|\theta) d\tau(\theta)$ è formalmente identica alla definizione di $\mu([\chi])$ come integrale della funzione caratteristica di $[\chi]$ rispetto a una misura su uno spazio di modelli. La differenza è che Carnap lavora con modelli parametrici semplici, mentre Gaifman lavora con modelli arbitrari. Il teorema di rappresentazione di Carnap–Stegmüller, secondo cui ogni funzione di conferma regolare e simmetrica è rappresentabile come probabilità bayesiana rispetto a una misura a priori, è un caso particolare del teorema di Scott: ogni misura logica è rappresentabile come misura classica su uno spazio di probabilità. La duplicazione è solo apparente: Carnap è Gaifman in un caso parametricamente semplice.

La costruzione di Keisler introduce quantificatori probabilistici $(P \times \geq r) \chi(x)$, che sembrano aggiungere qualcosa di radicalmente nuovo. In realtà, Keisler non fa altro che rendere esplicita la

lettura misura-teoretica dei quantificatori già implicita in Gaifman. La differenza è tecnica: Gaifman interpreta $\forall x \chi(x)$ come limite di frequenze, Keisler interpreta $(P x \geq r) \chi(x)$ come misura dell'insieme degli elementi che soddisfano χ . La costruzione tramite misure di Loeb non è una duplicazione della costruzione tramite ultrapotenze di Gaifman: è la stessa tecnica applicata a un dominio interno iperfinito invece che allo spazio dei modelli. Il teorema di completezza di Keisler è un'estensione del teorema di Gaifman: ogni teoria consistente con quantificatori probabilistici è soddisfatta in una struttura misurata. La duplicazione è solo apparente: Keisler è Gaifman con quantificatori probabilistici espliciti.

La teoria di Hoover sulle misure scambiabili sembra ancora più distante, ma anche qui la struttura è la stessa. Una misura scambiabile su strutture infinite è rappresentabile come miscela di misure prodotto, esattamente come una misura scambiabile su sequenze è rappresentabile come miscela di misure Bernoulli (teorema di de Finetti). Il teorema di Hoover è un teorema di rappresentazione misura-teoretica per modelli logici, e la sua forma astratta è identica a quella di Scott–Krauss: ogni misura logica scambiabile è rappresentabile come integrale di misure strutturalmente semplici. La duplicazione è solo apparente: Hoover è Scott–Krauss con simmetria.

A questo punto diventa naturale formulare un teorema di unificazione, che nella letteratura non compare come tale ma che sintetizza l'intero quadro.

Teorema (Fagin–Gaifman–Scott, forma unificata).

Sia L un linguaggio del primo ordine numerabile. Le seguenti condizioni sono equivalenti per una funzione μ definita sulle formule chiuse di L :

1. μ è una misura logica nel senso di Gaifman.
2. μ è rappresentabile come probabilità classica su uno spazio di modelli Q .
3. μ è rappresentabile come probabilità classica su uno spazio di teorie complete $S(L)$.
4. μ è rappresentabile come integrale rispetto a una misura a priori su uno spazio parametrico (Carnap).
5. μ è estendibile a una struttura con quantificatori probabilistici soddisfacibile in un modello di Loeb (Keisler).
6. μ è rappresentabile come miscela di misure scambiabili su strutture infinite (Hoover).

La dimostrazione non è un singolo argomento, ma una catena di equivalenze: Gaifman implica Scott tramite Stone duality, Scott implica Carnap tramite rappresentazione parametrica, Carnap implica Keisler tramite estensione dei quantificatori, Keisler implica Hoover tramite ultrapotenze e simmetria, Hoover implica Gaifman tramite de Finetti generalizzato. La duplicazione è solo apparente: tutte queste teorie sono facce della stessa struttura.

Sia $LT(L)$ l'algebra di Lindenbaum–Tarski delle formule chiuse di un linguaggio del primo ordine L , e sia $\mu : LT(L) \rightarrow [0,1]$ una misura logica σ -additiva. Il teorema unificato mostra che tutte le principali costruzioni della logica probabilistica — Gaifman, Carnap, Scott–Krauss, Keisler, Hoover — producono, a meno di isomorfismo di measure algebras, la stessa struttura di misura.

Teorema (Faggin–Gaifman–Scott, forma unificata). Versione più tecnica

Per una misura logica μ su $LT(L)$ sono equivalenti le seguenti condizioni:

1. Esiste una misura di Gaifman μ_G definita sullo spazio dei modelli Q tale che

$$\mu([\varphi]) = \mu_G([\varphi]) \text{ per ogni formula chiusa } \varphi.$$

2. Esiste uno spazio di probabilità standard $(\Omega, \mathcal{P}, P)$ e una famiglia di variabili casuali $\{1_{[\varphi]}\}$ tali che

$$\mu([\varphi]) = P(1_{[\varphi]} = 1).$$

3. Esiste uno spazio parametrico Θ e una misura a priori τ tali che per ogni formula chiusa φ esiste una funzione $f_{\varphi} : \Theta \rightarrow \{0,1\}$ con

$$\mu([\varphi]) = \int f_{\varphi}(\theta) d\tau(\theta).$$

(Questo generalizza il caso di Carnap.)

4. Esiste una struttura di Loeb (D^*, I^*, μ_L) ottenuta tramite ultrapotenze interne e standard part tale che

$$\mu([\varphi]) = \mu_L(\{a \in D^* : M^* \models \varphi(a)\}).$$

5. Se μ è scambiabile, esiste una rappresentazione alla Hoover del tipo

$$\mu([\varphi]) = \int 1_{[\varphi]}(F(U_1, \dots, U_k)) dU,$$

dove F è una funzione simmetrica e $\{U_i\}$ è una famiglia di variabili uniformi indipendenti.

Schema di dimostrazione

$(1 \Rightarrow 2)$

Si applica la dualità di Stone all'algebra $LT(L)$ e si usa il teorema di rappresentazione delle measure algebras per ottenere uno spazio standard $(\Omega, \mathcal{P}, P)$ isomorfo alla measure algebra generata da μ .

$(2 \Rightarrow 3)$

Si usa la disintegrazione della misura (teorema di Rokhlin) per rappresentare P come miscela di misure condizionate su un parametro $\theta \in \Theta$.

$(3 \Rightarrow 4)$

Si costruisce una catena di strutture finite che approssimano le distribuzioni $f_{\varphi}(\theta)$ e si forma l'ultrapotenza interna $M^* = \prod M_i / U$.

La misura interna μ_{int} definita tramite conteggio interno viene trasformata nella misura di Loeb μ_L tramite standard part.

(4 $\Rightarrow$ 1)

Si usa il lemma di Łoś per garantire che la valutazione interna delle formule in M^* rispetti la logica del primo ordine, e si definisce μ_G come pushforward di μ_L sugli insiemi $[\varphi]$.

(2 + simmetria $\Rightarrow$ 5)

Si applica il teorema di Aldous–Hoover–Kallenberg per rappresentare ogni misura scambiabile come miscela di misure prodotto definite da una funzione simmetrica F .

(5 $\Rightarrow$ 3)

Si interpreta la funzione simmetrica F come parametro θ e si ottiene una rappresentazione integrale del tipo

$$\mu([\varphi]) = \int f_{\varphi}(\theta) d\tau(\theta).$$

In tutti i casi, le strutture ottenute sono isomorfe come measure algebras: ciò significa che, pur essendo costruite in modi diversi (modelli finiti, ultrapotenze, misure di Loeb, spazi parametrici, rappresentazioni scambiabili), esse rappresentano la stessa misura logica su $LT(L)$.

Se ora volgiamo lo sguardo indietro, quello che emerge è che la “logica probabilistica” è meno una famiglia di logiche speciali e più un corridoio che attraversa l’intera architettura della logica e della probabilità: dalle misure su spazi di modelli (Gaifman) alle misure logiche (Scott-Krauss), dai quantificatori probabilistici e misure di Loeb (Keisler, Hoover) alle rappresentazioni scambiabili (Hoover, Aldous–Hoover), dal proposizionale probabilistico e politopi convessi (Fagin-Halpern-Megiddo, Nilsson) ai sistemi dinamici e logiche temporali probabilistiche, fino alle categorie di Markov, alla monade di Giry e ai calcoli probabilistici di ordine superiore. In ogni punto, gli stessi ingredienti ritornano: σ -algebre generate da formule o da eventi, misure finitamente additive estese a σ -additive, ultrafiltri e ultrapotenze per costruire modelli saturi, rappresentazioni come miscele di prodotti, invarianti di scambiabilità, operazioni monadiche che trasformano misure in misure.

Per rendere ancora più completo il quadro, dobbiamo spingerci su tre fronti: le connessioni con la teoria degli insiemi e la misurabilità “globale” (Solovay, insiemi di reali tutti misurabili e il ruolo dei grandi cardinali), le logiche infinitarie con quantificatori probabilistici e le leggi zero–uno su strutture casuali, infine la dimensione apprenditiva: semantica PAC, VC-dimension e limiti di apprendimento per frammenti di logica del primo ordine. E poi, ancora, la fascia filosofica: la tensione fra verità, probabilità, interpretazioni frequentiste e bayesiane, e l’idea che la probabilità sia il modo naturale di parlare di “verità parziale” in logica.

Partiamo dal legame con la teoria degli insiemi. Quando si parla di probabilità su spazi infiniti non è affatto banale chiedere che “tutti” i sottoinsiemi siano misurabili; sappiamo che, sotto l’assioma della scelta pieno (AC), esistono sottoinsiemi non misurabili di $[0,1]$: i classici insiemi di Vitali. Una via per “ripulire” il paesaggio è quella aperta da Solovay: partendo da $ZF + DC$ (assioma di dipendenza numerabile) e assumendo l’esistenza di un cardinale inaccessibile, è possibile costruire un modello in cui ogni sottoinsieme di $\mathbb{R}$ è Lebesgue-misurabile, ha la proprietà di Baire e il perfetto set property. Springer In tale modello, la probabilità (Lebesgue) assume un carattere “totale”: non ci sono più anomalie come insiemi di misura indefinita. Il costo è set-teorico: l’esistenza di un cardinale inaccessibile è un’ipotesi molto forte, e la rinuncia a pieno AC significa che non tutte le costruzioni insiemistiche possibili sono ammesse. Nel contesto delle logiche probabilistiche, queste considerazioni diventano rilevanti quando si vuole estendere la misurabilità a tutte le classi definibili

con certe risorse logiche: ad esempio, chiedere che tutti i sottoinsiemi definibili di $[0,1]$ da formule del secondo ordine siano misurabili può usare assiomi analoghi. Lo stesso discorso si riverbera nelle teorie delle probabilità “ben comportate”, in cui si desidera che ogni insieme definibile (per esempio ogni insieme boreliano, o più in generale ogni proiettivo) sia misurabile rispetto a una misura “canonica”.

Una delle tensioni che emergono qui è la seguente: la logica è, per costruzione, un motore generatore di sottoinsiemi definibili; la teoria della misura classica ci dice che non tutti i sottoinsiemi sono misurabili sotto AC; la teoria degli insiemi rafforzata con grandi cardinali dice: se desideri una nozione “universale” di probabilità, non basta la ZFC ordinaria, devi muoverti in un universo set-teorico più ricco. Così la logica probabilistica, nel suo anelito a trattare probabilisticamente “tutti” gli enunciati, si trova a toccare i limiti stessi della set-teoria.

La dimensione infinitaria delle logiche con probabilità è un altro punto in cui la combinazione logica-misura diventa delicata. Consideriamo una logica $L_{\omega_1\omega}$ arricchita con quantificatori probabilistici, in cui sono ammesse congiunzioni e disgiunzioni numerabili. Frasi di questa logica possono esprimere proprietà globali su strutture infinite in modo molto più potente del primo ordine. Ad esempio, in teoria dei grafi, si possono esprimere proprietà che codificano l'esistenza di catene infinite di condizioni locali; a maggior ragione, aggiungendo operatori probabilistici, si può parlare della probabilità che “in quasi tutti i sottoinsiemi finiti del dominio” una certa proprietà valga. In questo contesto trovano posto leggi zero-uno raffinate: per modelli random come il grafo casuale $G(n,p)$ con $n \rightarrow \infty$ e $0 < p < 1$ fisso, si dimostra che per ogni formula ϕ in un certo frammento di $L_{\omega_1\omega}$ arricchito con operatori probabilistici, la probabilità che $G(n,p) \models \phi$ tende a 0 o 1: non esistono limiti intermedi stabili. La struttura di questi teoremi è profondamente combinatoria: si usano giochi di Ehrenfeucht-Fraïssé adattati al contesto random, in cui due giocatori si sfidano a distinguere grafi sulla base di formule limitate, e si mostra che, con probabilità tendente a 1, tali giochi hanno sempre vincitore “trivialmente” determinato; di conseguenza, la teoria quasi-sicura (almost sure theory) del grafo casuale è completa per quel frammento: le formule hanno sempre una verità asintotica definita. Quando si aggiungono simboli di probabilità, la legge zero-uno acquista un colorito particolare: affermazioni che parlano di frequenze e densità si polarizzano, nel limite, in verità rigide.

Spingendo ancora oltre l'aspetto “globalità probabilistica”, si arriva alle logiche con quantificatori come “per quasi tutti” (in senso topologico o misurabile), un po' come nella stationary logic di Mundici: l'operatore $(P_x) \phi$ viene interpretato come “ $\phi(x)$ vale per quasi tutti gli x in un senso di misura” o talvolta “per tutti tranne che per un insieme di misura nulla”, e si studia la possibilità di eliminare questi quantificatori riducendo le formule a combinazioni finitarie più controllabili. Il risultato tipico è un teorema di eliminazione di quantificatori probabilistici per certi frammenti: ogni formula con (P_x) può essere trasformata in una formula equivalente che parla solo di misure di insiemi definibili, ossia di vincoli su numeri reali, rendendo possibile decidere (per certe classi di strutture) la soddisfacibilità di formule. Questo è un caso esemplare di come la logica probabilistica, per non sfuggire al controllo, si auto-imponga restrizioni sulle modalità di interazione fra quantificatori e misura.

Se spostiamo lo sguardo dal versante modellistico al versante computazionale e apprenditivo, la semantica PAC (Probably Approximately Correct) introdotta da Valiant entra in gioco come una “logica probabilistica della generalizzazione”. In PAC learning, un concetto è un sottoinsieme C di uno spazio X , e un'ipotesi è un altro sottoinsieme H di X ; c'è una distribuzione D su X sconosciuta, e si chiede che un algoritmo di apprendimento, ricevuti esempi i.i.d. da D etichettati secondo C , produca un'ipotesi H che approssimi C con alta probabilità: formalmente, per ogni $\epsilon, \delta > 0$, dopo un numero polinomiale in $1/\epsilon, 1/\delta$ di esempi, l'algoritmo produce H tale che con probabilità almeno $1-\delta$,

la misura (secondo D) del simmetrico $C \Delta H \leq \varepsilon$. In forma compatta: φ è (ε, δ) -valida in senso PAC se

$$\Pr_{\{\text{sample} \sim D^m\}}[\Pr_{\{x \sim D\}}[\varphi(x)] \geq 1 - \varepsilon] \geq 1 - \delta.$$

Tradotto in termini logici, una classe di formule è PAC-learnable se esiste un algoritmo che, da esempi di verità/falsità di formule sotto qualche distribuzione su strutture, produce con alta probabilità una formula che “si comporta” quasi come la formula target. I collegamenti con la logica del primo ordine emergono quando si considera, ad esempio, il problema di apprendere definizioni FO su strutture con un vincolo sulla profondità quantificazionale. Risultati in stile Kearns–Vazirani–Kushilevitz–Mansour mostrano che frammenti del primo ordine con profondità quantificazionale limitata e vocabolario limitato hanno dimensione VC finita: la dimensione VC di una classe di concetti C (qui l’insieme delle insiemi definibili da formule di quel frammento) è il massimo k per cui esiste un insieme di punti di X che viene “shatterato” da C, cioè per cui ogni etichettatura binaria è realizzata da qualche concetto in C. Se la dimensione VC è finita, la classe è PAC-learnable; se è infinita, no. La cosa notevole è che, per certe classi FO restritte (per esempio, formule con un numero fissato di quantificatori alternati e struttura topologica limitata), la dimensione VC è finita; per FO pieno, è tipicamente infinita. Così, la logica probabilistica si manifesta come un vincolo di complessità apprenditiva: aggiungere potenza espressiva al linguaggio significa quasi sempre “spacchettare” il dominio in un numero talmente grande di configurazioni separabili da rendere impossibile il controllo statistico con un numero pratico di esempi.

C’è poi la dimensione dell’apprendimento online con esperti, dove una teoria logica viene vista come un “esperto” che produce previsioni probabilistiche su eventi in sequenza. Al tempo t, l’esperto T (cioè una teoria T) produce una distribuzione p_t sugli eventi possibili, e poi subisce una perdita $\ell_t(p_t)$ che dipende dall’esito reale e dalla distribuzione proposta. Il regret accumulato fino a tempo T è

$$R_T = \sum_{t=1}^T \ell_t(p_t) - \min_{T' \in \mathcal{T}} \sum_{t=1}^T \ell_t(p_{t'}),$$

dove $\mathcal{T}$ è una famiglia (anche infinita) di teorie candidate. Risultati classici di Cesa-Bianchi & Lugosi mostrano che, sotto convessità di ℓ_t e qualche vincolo sulla complessità di $\mathcal{T}$, esistono algoritmi (come l’aggregazione esponenziale) che garantiscono $R_T = O(\sqrt{T \log |\mathcal{T}|})$ per famiglie finite $\mathcal{T}$ e più in generale $R_T = o(T)$ per famiglie infinite con “complessità effettiva” finita (per esempio, entropia metrico). In chiave logica, questo significa che una famiglia infinita di teorie probabilistiche può essere “seguita” con regret sublineare se la sua struttura è sufficientemente regolare, e dunque che, in un certo senso, anche il problema di scegliere fra teorie logiche probabilistiche concorrenti è soggetto a vincoli di apprendimento che dipendono dalla dimensione della famiglia. Ancora una volta, la probabilità funziona da ponte fra struttura logica (famiglia di teorie) e limiti operativi (regret, complessità apprenditiva).

Un altro fronte che va approfondito è quello delle probabilità infinitesimali e delle funzioni di Popper. L’idea di introdurre probabilità infinitesimali in logica nasce dall’esigenza di trattare eventi “impossibili ma non proprio zero”: nel classico calcolo delle probabilità, un evento di probabilità esattamente zero è tipicamente considerato “impossibile” in senso numerico, ma non necessariamente in senso logico (accade spesso in modelli continui). Le funzioni di Popper cercano di definire una probabilità condizionata $P(A|B)$ direttamente come primitiva, anche quando $P(B) = 0$, e richiedono che $P(\cdot|B)$ sia una probabilità finita su un’anello di insiemi, e che $P(A \wedge B) = P(A|B) P(B)$, dove $P(B)$ può essere un valore infinitesimale. Per modellare questo, si lavora spesso con campi di numeri non standard: la probabilità P è una funzione che prende valori in $^*[0,1]$, e gli eventi di probabilità “quasi zero” sono quelli la cui probabilità è infinitesimale ma non esattamente 0. Da un punto di vista logico,

le funzioni di Popper permettono di definire sistemi di condizionali “per condizionamento su eventi di probabilità zero”: formule tipo $\varphi > \psi$ (“se φ allora ψ ” in senso condizionale) vengono interpretate come asserzioni su $P(\psi \mid \varphi)$. Questo porta a logiche condizionali probabilistiche molto raffinate, in cui si possono distinguere enunciati come “dati tutti i casi di misura zero in cui φ è vera, ψ è quasi sempre vera” dai casi in cui φ rende ψ solo “spesso” vera ma non quasi sempre. In letteratura, si coglie qui un legame con la logica delle credenze, delle revisioni (AGM) e dei ranking functions, in cui la probabilità infinitesimale viene vista come un raffinamento della nozione di plausibilità.

Infine, per completare il quadro fondazionale, vale la pena soffermarsi sulla dimensione filosofica di “probabilità e verità”. Una linea classica, che risale a Tarski, identifica la verità di una formula con la sua soddisfazione in un modello; le logiche probabilistiche introducono una “scala” di valori intermedi: il valore di verità di φ non è più semplicemente vero o falso, ma una probabilità in $[0,1]$. Tuttavia, questa probabilità non è un generico grado di verità fuzzy: è sempre una probabilità in senso classico, definita rispetto a uno spazio di possibilità (modelli, strutture, stati, cammini), che si restringe a $\{0,1\}$ quando la formula è o quasi-sicuramente vera o quasi-sicuramente falsa. L’interpretazione frequentista vede questa probabilità come limite di frequenza relativa di strutture (Gaifman, logiche zero–uno); l’interpretazione bayesiana la vede come grado di credenza razionalmente aggiornato (Carnap, de Finetti, Gaifman sui gradi di credenza in arithmeticfitelson.org+1); l’interpretazione logico–modale la vede come modalità “in quasi tutti i mondi” ($P \times \geq r \varphi(x)$ visto come un operatore modale quantitativo). In tutti questi casi, il concetto di verità viene destrutturato in “verità quasi sicura”, “verità con grado r ”, “verità nella misura-uno dei mondi”, ma il nucleo rimane: la probabilità non sostituisce la verità, ne modula il campo di applicazione.

Questa idea si riflette nell’uso della probabilità in teorie della verità parziale e dei paradossi. Se si introduce una nozione di verità probabilistica, $T(\varphi) = p$, si può tentare di gestire enunciati auto-referenziali tipo il mentitore (“questa frase è falsa”) assegnando loro una distribuzione di probabilità sulla verità piuttosto che un valore binario. Le speranze sono di “diffondere” la contraddizione su una distribuzione non degenerata, evitando l’esplosione. Tuttavia, risultati negativi mostrano che molte versioni del paradosso del mentitore si ripresentano in forma probabilistica: condizioni naturali sulla funzione T e sulla coerenza probabilistica conducono ancora a incoerenze o a cicli di aggiornamento infiniti. Il “ritorno del mentitore in veste probabilistica” mostra che il mero passaggio alla probabilità non immunizza la logica dai paradossi di auto-riferimento: occorre intervenire a livello di schema di verità, o di strutture (ad esempio usando gerarchie di livelli o logiche non classiche).

Se riassumiamo questo percorso, ma senza fornire un riassunto conclusivo, il filo è chiaro: ogni volta che la logica tenta di dire qualcosa di globale – su tutti i modelli, su quasi tutti i modelli, sulla struttura asintotica di modelli grandi, sulla generalizzabilità da esempi, sulla revisione delle credenze – la probabilità entra come linguaggio naturale di questa globalità. La logica probabilistica non è una logica “in più”, ma il modo in cui la logica esprime le proprie ambizioni estensive: misurare l’insieme dei mondi, disciplinare il limite delle strutture, quantificare la robustezza di una teoria rispetto a dati e incertezze.

L’idea che la probabilità possa essere intrecciata alla logica non come un’aggiunta esterna, ma come un’estensione naturale della semantica, si manifesta in modo pieno quando si osserva che ogni formula logica definisce un insieme di modelli, e che questi insiemi possono essere organizzati in una σ -algebra sulla quale definire una misura. La costruzione di Gaifman, che abbiamo già delineato, è il primo passo: si prende un linguaggio del primo ordine L numerabile, si considera l’insieme Q di tutte le strutture per L , e si genera una σ -algebra Σ_L dagli insiemi $[\varphi] = \{ M \in Q \mid M \models \varphi \}$.

Una volta fissato che ogni formula chiusa ϕ in un linguaggio del primo ordine L definisce un insieme di modelli $[\phi]$ dentro uno spazio Q di strutture, e che la collezione di tali insiemi genera una σ -algebra Σ_L su cui una misura μ può essere definita, diventa naturale chiedere sotto quali condizioni questa misura rappresenti davvero una “probabilità logica” coerente con i connettivi e i quantificatori. Le condizioni elementari sono chiare: $\mu([\top]) = 1$, $\mu([\perp]) = 0$, $\mu([\phi \vee \psi]) = \mu([\phi]) + \mu([\psi]) - \mu([\phi \wedge \psi])$, σ -additività su famiglie numerabili di eventi disgiunti. Per i quantificatori, come abbiamo visto, una via consiste nel legare $\mu([\forall x \psi(x)])$ a limiti di eventi finiti che coinvolgono costanti nuove: si introducono costanti $c_1, c_2, \dots$ e si chiede che

$$\mu([\forall x \psi(x)]) = \lim_{n \rightarrow \infty} \mu([\psi(c_1) \wedge \dots \wedge \psi(c_n)]),$$

quando il limite ha senso ed è ben definito nel modello. Questo collega il quantificatore universale a una frequenza asintotica di soddisfacimento da parte delle costanti enumerate. Un punto fine è che, in generale, questo limite potrebbe non esistere; le condizioni di regolarità su μ e sulla scelta delle costanti servono proprio a garantire l'esistenza e l'unicità del limite, almeno per un'ampia classe di formule.

L'uso di ultraprodotti e di tecniche non standard nella dimostrazione di esistenza di una tale μ risponde a una difficoltà fondamentale: è molto complicato controllare direttamente limiti di frequenze su domini infiniti standard; è più semplice costruire, tramite ultrapotenze, domini iperfiniti in un universo non standard in cui le frequenze diventano semplici rapporti di numeri interi interni. La misura interna μ_{int} , definita come $\mu_{\text{int}}(A) = |A|_{\text{int}} / |D|_{\text{int}}$ per insiemi interni $A \subseteq D$, è finitamente additiva per costruzione. Il passaggio alla misura esterna μ_L mediante la standard part permette di ottenere una misura σ -additiva su una σ -algebra Σ_L su D^* , con $\mu_L(A) = \text{st}(\mu_{\text{int}}(A))$ per insiemi interni A . Da quel momento, quantificatori come $(\exists x \geq r) \phi(x)$ possono essere interpretati rispetto a μ_L , e la logica con quantificatori probabilistici trova un modello concreto.

Il teorema di completezza di Keisler, nella sua forma concettuale, afferma che il sistema assiomatico che governa questi quantificatori probabilistici è esattamente calibrato su questo tipo di semantica: una formula è dimostrabile se e solo se è vera in tutte le strutture di Loeb. Per mostrare la direzione “se una formula non è dimostrabile allora ha un modello di Loeb che la falsifica”, si parte da un insieme di formule che nega la formula in questione e si mostra che è consistente con il sistema assiomatico; poi, con un argomento alla Henkin, si completa questa teoria e si costruisce una catena di modelli finiti che approssimano progressivamente la teoria. Gli ultrafiltri non principali e l'ultraprodotto forniscono il modello non standard M , nel quale la misura interna sui sottoinsiemi del dominio rispecchia, a livello di frequenze, i vincoli imposti dai quantificatori probabilistici sulle formule. La misura di Loeb μ_L che si ottiene ha allora la proprietà desiderata: per ogni formula $(\exists x \geq r) \phi(x)$ nella teoria, $\mu_L(\{a \mid M \models \phi(a)\}) \geq r$; per ogni negazione coerente con la teoria, il vincolo viene rispettato nella direzione opposta. Verificare questa corrispondenza richiede di usare il lemma di Łoś per il trasferimento di verità nelle ultrapotenze: se una formula ϕ con parametri è vera in quasi tutti gli M_i , allora è vera in M ; e viceversa, se è falsa nella maggior parte, sarà falsa anche in M . Le frequenze sulle sequenze finite M_i si trasferiscono in μ_{int} , e quindi in μ_L , come vincoli numerici reali.

Il fatto che le misure di Loeb trasformino misure interne finitamente additive in misure esterne σ -additive è cruciale per mantenere la piena teoria della misura: integrali, limiti monotoni, teorema della convergenza dominata, ecc., diventano disponibili nel modello non standard. Questo giustifica l'uso dei quantificatori probabilistici come veri quantificatori di misura, anziché come meri gadget finitamente additivi. Una logica che pretendesse di lavorare con misure solo finitamente additive sul dominio perderebbe una buona parte delle proprietà strutturali desiderate (come il teorema di Radon–

Nikodým, la disintegrazione, la possibilità di definire condizionali regolari), che diventano invece recuperabili via Loeb.

Un discorso parallelo si può fare per le misure su spazi di strutture. Qui il ruolo dei numeri non standard è meno centrale; ciò che conta è l'idea di scambiabilità e di rappresentazione. Una misura μ su un insieme di strutture di un linguaggio L con dominio numerabile è scambiabile se è invariante sotto permutazioni del dominio: se π è una permutazione di $\mathbb{N}$, e M^π è la struttura ottenuta rinominando tutti gli elementi del dominio di M secondo π , allora $\mu(A) = \mu(\{ M^\pi \mid M \in A \})$ per ogni insieme misurabile A . Questa è la generalizzazione naturale della scambiabilità di de Finetti per sequenze: l'ordine delle etichette degli elementi del dominio è irrilevante, conta solo la struttura indotta.

Nel caso più semplice di un linguaggio monadico con un solo predicato R , una struttura M è solo un sottoinsieme di $\mathbb{N}$: l'insieme degli indici i per cui $R(i)$ vale. Una misura μ su queste strutture scambiabile è l'equivalente di una misura su $\{0,1\}^\mathbb{N}$ invariante sotto permutazioni delle coordinate. Il teorema di de Finetti garantisce che ogni tale μ può essere rappresentata come una miscela di misure prodotto i.i.d.:

$$\mu(B) = \int 0^I 1 \mu_\theta(B) d\pi(\theta),$$

dove μ_θ è la misura prodotto corrispondente a variabili indipendenti Bernoulli(θ) e π è una misura a priori su $[0,1]$. Questo significa che la struttura casuale definita da μ è, condizionata a un parametro θ , un modello in cui ogni elemento soddisfa R indipendentemente con probabilità θ . In linguaggio logico: una teoria probabilistica scambiabile con un predicato monadico è equivalente a scegliere prima un valore di $\theta \in [0,1]$ secondo π e poi, dato θ , scegliere un modello i.i.d. in cui la verità di $R(x)$ è indipendente per ogni x e ha probabilità θ .

Per linguaggi con simboli di relazione di arità ≥ 2 , l'analogo è molto più complicato: per una relazione binaria $R(x,y)$, ad esempio, una struttura è un grafo (orientato o meno) su $\mathbb{N}$, e una misura scambiabile è una misura su grafi infiniti invariante sotto permutazioni dei vertici. Il teorema di Aldous–Hoover generalizza de Finetti a matrici scambiabili $(X_{ij})_{i,j \in \mathbb{N}}$. Per semplicità, prendiamo $X_{ij} \in \{0,1\}$, con $X_{ij} = X_{ji}$ e $X_{ii} = 0$, in modo da rappresentare grafi semplici non orientati. L'ipotesi di scambiabilità richiede che, per ogni permutazione π di $\mathbb{N}$, il processo (X_{ij}) abbia la stessa legge di $(X_{\pi(i)\pi(j)})$. Il teorema dice che esistono variabili casuali $U \sim \text{Unif}[0,1]$, $U_i \sim \text{i.i.d. Unif}[0,1]$ per $i \in \mathbb{N}$, e $W_{ij} \sim \text{i.i.d. Unif}[0,1]$ per $i < j$ (indipendenti tra loro e da U e dagli U_i), e una funzione misurabile simmetrica $f: [0,1]^3 \rightarrow \{0,1\}$ tale che

$$X_{ij} = f(U, U_i, U_j, W_{ij}),$$

dove per semplicità possiamo assorbire U nel dominio di f o trattarlo come parametro globale. In pratica, la struttura casuale del grafo è generata da un parametro globale U , da etichette latenti U_i per ogni vertice, e dal rumore W_{ij} per ogni coppia: condizionatamente alla collezione $(U, (U_i))$, gli archi X_{ij} sono indipendenti. Dal punto di vista logico, questo significa che esiste una “rappresentazione fattorizzata” della misura scambiabile μ su grafi: μ è uguale a $\int \mu_\theta d\nu(\theta)$, dove θ incapsula la funzione f e la distribuzione di (U, U_i) , e μ_θ è una misura prodotto condizionata a θ . Hoover ha generalizzato questo quadro a linguaggi più generali, mostrando che misure scambiabili su strutture per linguaggi arbitrari (con relazioni di arità finita) possono sempre essere viste come miscele di misure in cui i fatti atomici sono condizionatamente indipendenti dato un opportuno parametro.

Il ruolo della definibilità in questo contesto è quello di assicurare che le proprietà probabilistiche rilevanti delle strutture possano essere espresse nel linguaggio L o in sue estensioni moderate. Una misura μ su strutture di L è definibile in L se per ogni formula $\varphi(x_1, \dots, x_n)$ e ogni $\varepsilon > 0$ esiste una formula priva di quantificatori $\psi(x_1, \dots, x_n)$ tale che la differenza tra gli insiemi di n -uple che soddisfano φ e ψ è, in un senso probabilistico, piccola. Formalizzando: per “quasi tutte” le strutture M rispetto a μ , vale che la misura (secondo una certa misura canonica sul dominio, ad esempio la misura di conteggio normalizzata su insiemi finiti) dell’insieme

$$\{ a \in D^n \mid M \models \varphi(a) \leftrightarrow \psi(a) \}$$

è almeno $1 - \varepsilon$. Questa approssimazione quantificatore-free è essenziale per portare la struttura complessa del linguaggio quantificato dentro la componente prodotto del modello: ciò che i quantificatori aggiungono può essere quasi sempre assorbito nella scelta delle etichette latenti e nella funzione f .

Tutto questo ha un riflesso naturale nelle logiche infinitarie con probabilità. In una logica $L_{\omega_1\omega}$ si ammettono congiunzioni e disgiunzioni numerabili, ma quantificatori solo finiti. Se si aggiungono quantificatori probabilistici, si possono scrivere formule che combinano limiti su catene numerabili di enunciati e vincoli sulla loro misura. Ad esempio, una formula può affermare che “per ogni $\varepsilon > 0$, esiste un N tale che per tutti $n \geq N$ la misura dell’insieme degli x per cui $\varphi_n(x)$ è vera è $\geq 1 - \varepsilon$ ”, dove (φ_n) è una sequenza di formule. Questo tipo di espressione consente di parlare di convergenze in misura e di proprietà asintotiche di strutture infinite, collegando direttamente logica e analisi probabilistica.

Le leggi zero–uno sono uno dei luoghi in cui questa connessione emerge con forza. Per la logica del primo ordine, una famosa legge zero–uno afferma che, per il grafo casuale $G(n, p)$ con $n \rightarrow \infty$ e $0 < p < 1$ fisso, per ogni formula FO φ in linguaggio dei grafi, la probabilità che $G(n, p) \models \varphi$ tende a 0 o a 1. In altre parole, la teoria quasi-sicura del grafo casuale è completa: per ogni φ , o φ è quasi sicuramente vera o quasi sicuramente falsa nel limite. La dimostrazione usa il metodo delle partizioni regolari e argomenti di Ehrenfeucht–Fraïssé: si mostra che per ogni k , con probabilità tendente a 1, i grafi grandi sono indistinguibili da un certo grafico limite rispetto a giochi a k mosse; di conseguenza, qualunque formula con profondità quantificazionale $\leq k$ ha valore di verità stabilizzato. Quando si aggiungono operatori probabilistici, si può ottenere, in certe classi di strutture random, leggi zero–uno raffinate in cui le affermazioni su misure di insiemi definibili hanno anch’esse limiti 0 o 1, oppure, in casi più delicati, limiti che dipendono solo dai parametri p e da invarianti strutturali globali.

Il passaggio alle probabilità infinitesimali, tramite funzioni di Popper, modifica profondamente la geometria semantica. Una funzione di Popper P è una funzione che assegna a coppie di insiemi (A, B) con $B \neq \emptyset$ un numero $P(A|B)$ in un campo ordinato esteso, spesso un campo non standard contenente infinitesimi. Le condizioni che P deve soddisfare mimano le proprietà della probabilità condizionata: per ogni B , la funzione $A \mapsto P(A|B)$ è finitamente additiva, $P(B|B) = 1$, $P(A \wedge B | B) = P(A|B)$, e $P(A \wedge C | B) = P(A|B) P(C|A \wedge B)$ (una forma di catena). Il punto è che non si richiede che esista una misura incondizionata $P(\cdot)$ tale che $P(A|B) = P(A \wedge B) / P(B)$; anzi, si permette che $P(B)$ sia infinitesimale. In questo modo, eventi “di probabilità zero” in senso standard possono essere trattati con probabilità infinitesimali non nulle, e condizionali come $P(A|B)$ possono essere definiti anche quando B ha misura standard zero ma misura non standard infinitesimale. La combinazione di funzioni di Popper con logiche condizionali consente di trattare enunciati del tipo “se φ , allora ψ ” come condizioni su $P(\psi|\varphi)$: $\varphi > \psi$ è vero se $P(\psi|\varphi)$ è “quasi 1”, cioè 1 meno un infinitesimo. Questo produce sistemi in cui si distingue finemente fra casi in cui ψ è sempre vero quando φ è vero, casi in cui ψ è vero “quasi sempre” (ma con eccezioni di misura infinitesimale), e casi in cui ψ è solo più probabile di $\neg\psi$ condizionato a φ .

Dal punto di vista della teoria della misura classica, le probabilità infinitesimali sono estranee; ma dal punto di vista della non standard analysis e delle misure di Loeb, sono naturali: ogni misura di Loeb nasce da una misura interna che produce valori in ${}^*[0,1]$, e la parte standard è un collasso di informazioni infinitesimali. L'idea di funzioni di Popper può essere vista come un tentativo di "restaurare" quella parte infinitesimale come oggetto semantico di primo piano. C'è un collegamento interessante con la logica delle credenze qualitative: le ranking functions di Spohn, che assegnano a proposizioni numeri ordinali che rappresentano il loro grado di "incredibilità", possono essere viste come $\log(1/\varepsilon)$ -trasformate di probabilità infinitesimali; in questo senso, logiche condizionali qualitative e funzioni di Popper sono due facce della stessa medaglia.

L'altra grande linea che abbiamo solo abbozzato è quella set-teorica: il ruolo dei grandi cardinali e dei modelli in cui ogni insieme di reali è misurabile. Solovay mostra che, assumendo l'esistenza di un cardinale inaccessibile κ , si può costruire un modello di $ZF + DC$ in cui ogni sottoinsieme di $\mathbb{R}$ è Lebesgue misurabile, ha la proprietà di Baire e la perfect set property. In questo modello, la nozione di probabilità su $[0,1]$ (rispetto alla misura di Lebesgue) è universale: non ci sono controesempi patologici come gli insiemi di Vitali. La costruzione usa un forcing raffinato e la cancellazione dell'assioma della scelta piena, mantenendo però la dipendenza numerabile. Dal punto di vista della logica probabilistica, questo modello è affascinante perché rende possibile un'idea estrema: una logica in cui "ogni" proprietà di insiemi di reali definibile in un linguaggio abbastanza ampio è misurabile e ha probabilità definita. Tuttavia, il prezzo è alto: l'esistenza di cardinali inaccessibili è indipendente da ZFC e viene spesso considerata una ipotesi forte; d'altra parte, molte teorie moderne della misura (supercompatte, leggi zero-uno per classi molto ampie di insiemi) richiedono ipotesi simili. Così, la pura aspirazione logica a una probabilità "totale" su un universo di discorso ricco ci spinge inevitabilmente nel territorio dei grandi cardinali.

Quando si passa al livello dell'apprendimento e della complessità, la probabilità assume un ruolo operativo. La semantica PAC, come dicevamo, è un modello per l'apprendimento di concetti (e quindi di formule) a partire da esempi casuali. Sia X uno spazio di oggetti (per esempio, strutture finite, o tuple di elementi di una struttura), e C una classe di concetti, cioè una famiglia di sottoinsiemi di X . Per ogni $c \in C$, c rappresenta un possibile "vero concetto" da apprendere. Un algoritmo di apprendimento osserva esempi $x_1, \dots, x_m$ i.i.d. da una distribuzione D su X , etichettati da $y_i = 1$ se $x_i \in c$, 0 altrimenti, e produce un'ipotesi $h \in C'$ (spesso C stessa) tale che, con alta probabilità sulla scelta degli esempi, l'errore generalizzato $\text{err}_D(h, c) = \Pr_{x \sim D}[h(x) \neq c(x)]$ è piccolo. Una classe C è PAC-learnable se esiste un algoritmo che, per ogni $\varepsilon, \delta > 0$, con $m = \text{poly}(1/\varepsilon, 1/\delta)$ esempi, garantisce $\text{err}_D(h, c) \leq \varepsilon$ con probabilità $\geq 1 - \delta$ per ogni $c \in C$ e ogni D .

La teoria della dimensione VC dà una caratterizzazione quasi completa di quando una classe C di insiemi è PAC-learnable: se $\text{VCdim}(C)$ è finito, allora C è PAC-learnable; se è infinito, no (fatto salvo qualche dettaglio tecnico su complessità computazionale). Applicata alle classi definibili logicamente, questa teoria si traduce in enunciati del genere: la famiglia di insiemi definibili da formule di un certo frammento logico (per esempio, formule quantificatore-libere su una struttura o formule FO con un numero fissato di quantificatori e arietà limitata) ha dimensione VC finita, quindi è PAC-learnable; invece, la famiglia di tutti i sottoinsiemi definibili da formule FO piene ha dimensione VC infinita, quindi non è PAC-learnable in generale. Questo collega potente concetti: il limite espressivo di un frammento logico determina il suo comportamento apprenditivo probabilistico. In altre parole, la logica, se vista come generatore di classi di concetti, incontra inevitabilmente le barriere poste dalla statistica: se il linguaggio è troppo potente, la sua capacità di "frammentare" X in regioni distinte rende impossibile generalizzare da campioni finiti.

L'apprendimento online con esperti aggiunge un'altra dimensione: invece di un solo concetto vero, abbiamo una famiglia di teorie o modelli che fanno previsioni probabilistiche sui dati in sequenza. A

ogni istante t , l'algoritmo osserva il contesto x_t , consulta una famiglia di esperti (teorie probabilistiche T_i) ognuno dei quali propone una distribuzione $p_t^{(i)}$ sull'esito y_t , e poi sceglie una distribuzione aggregata p_t . Dopo aver visto il vero esito y_t , l'algoritmo e gli esperti subiscono una perdita $\ell_t(p_t)$ e $\ell_t(p_t^{(i)})$. Il regret rispetto al miglior esperto i^* è

$$R_T = \sum_{t=1}^T \ell_t(p_t) - \sum_{t=1}^T \ell_t(p_t^{(i^*)}).$$

Le garanzie del tipo $R_T = O(\sqrt{T})$ implicano che, in media, l'algoritmo fa quasi bene quanto il miglior esperto, a posteriori, nella famiglia. Se gli esperti sono teorie probabilistiche logiche (per esempio, modelli bayesiani su strutture FO, o sistemi di regole probabilistiche), la dinamica di apprendimento online fornisce un criterio operativo per confrontare teorie: non più solo in termini di verità logica, ma in termini di prestazioni predittive cumulative. Dal punto di vista concettuale, questo disegna un ponte fra logica probabilistica e epistemologia bayesiana dinamica: non basta avere una teoria probabilistica coerente, occorre valutarne la capacità di reggere nel tempo sotto aggiornamenti ripetuti; le misure logiche di Gaifman, le funzioni di conferma di Carnap, le rappresentazioni di Hoover diventano ingredienti di modelli di previsione che, a loro volta, vengono sottoposti a selezione attraverso schemi di aggregazione.

Le fondazioni categoriali completano il quadro concettuale. La monade di Giry e le categorie di Markov forniscono un linguaggio astratto in cui la probabilità è codificata come struttura monadica su una categoria di base (ad esempio, la categoria degli spazi misurabili). Un object X rappresenta un tipo di stato; un morfismo $f: X \rightarrow G(Y)$, dove G è il funtore che manda X nell'insieme delle misure di probabilità su X , rappresenta un canale probabilistico che, dato un input in X , produce una distribuzione su Y . Composte f e $g: Y \rightarrow G(Z)$ tramite la moltiplicazione della monade $\mu: G(G(Z)) \rightarrow G(Z)$, otteniamo un canale $X \rightarrow G(Z)$ che rappresenta l'effetto combinato di due trasformazioni stocastiche. In termini di linguaggi probabilistici, questo è esattamente ciò che fa un programma probabilistico: compone istruzioni di campionamento (canali elementari) in strutture più complesse.

Le categorie di Markov astraggono ulteriormente: una Markov category è una categoria monoidale simmetrica in cui ogni oggetto X ha un comonoide (copia e cancellazione) che rappresenta la duplicazione e la marginalizzazione, e in cui i morfismi $X \rightarrow Y$ sono interpretati come mappe probabilistiche. Risultati recenti mostrano che molte proprietà classiche della probabilità (indipendenza, condizionamento, Bayes, sufficienza) possono essere formulate internamente in termini puramente categoriali, senza nominare misure o integrali. In questo linguaggio, la logica probabilistica può essere vista come la teoria delle formule invarianti sotto certe classi di morfismi; ad esempio, la scambiabilità diventa l'invarianza di un morfismo sotto l'azione di un gruppo di permutazioni, e il teorema di de Finetti si traduce in una factorization property unica di certi morfismi simmetrici.

Dal punto di vista logico, questo suggerisce un'estensione: la logica non è solo una teoria di relazioni tra insiemi e funzioni, ma può essere riformulata come una teoria di invarianti categoriali: enunciati che sono preservati sotto functor probabilistici, trasformazioni monoidali, ecc. Questo apre la strada a logiche probabilistiche "interne" a categorie di Markov, in cui si possono formulare enunciati su morfismi e misure direttamente nel linguaggio categoriale. L'analogo categorico della misura di Loeb, dei quantificatori probabilistici, della scambiabilità, delle miscele di prodotti, tutto può essere espresso in termini di oggetti, morfismi, universal property, e equazioni diagrammatiche.

Sullo sfondo di tutto questo rimane la questione filosofica: che cosa è, davvero, "logica probabilistica"? Non è solo un insieme di sistemi formali in cui compaiono simboli come $\Pr(\phi) \geq r$. È piuttosto l'insieme dei modi in cui la logica, come disciplina del possibile, si allea con la probabilità, come disciplina del misurabile. Quando diciamo che una teoria T ha probabilità 1 rispetto a una

misura μ su $S(L)$, stiamo affermando qualcosa di diverso da “ T è vera in tutti i modelli”: stiamo dicendo che, nel paesaggio delle strutture, le eccezioni sono di misura nulla. Quando affermiamo che un enunciato ϕ è PAC-valido, stiamo dicendo che ϕ è “quasi sempre” vero sotto tutte le distribuzioni che rispettano certi vincoli. Quando introduciamo un quantificatore ($P x \geq r$), stiamo raffinando $\forall$ in “per la maggior parte” e $\exists$ in “per una frazione non nulla”. Quando usiamo probabilità infinitesimali, stiamo separando l’impossibile logico dall’evento di misura zero, distinguendo fra ciò che non può accadere in nessun modello e ciò che accade, ma solo in un set di misura trascurabile.

In questo senso, la probabilità non sostituisce la verità, ma la stratifica: accanto a vera/ falsa, compaiono livelli come “quasi sicuramente vera”, “vera con grado r ”, “coerente ma di misura nulla”, “coerente e generica”. E la logica probabilistica è, nel profondo, la disciplina che studia come queste stratificazioni si inseriscono nei sistemi formali e nei modelli matematici: come si costruiscono misure su modelli, come si rappresentano misure logiche, come si introducono quantificatori di misura, come si caratterizzano misure scambiabili, come si valutano enunciati su strutture random, come si apprendono teorie da dati rumorosi.

Consideriamo ora le logiche proposizionali probabilistiche. Partiamo da un insieme finito di proposizioni atomiche $p_1, \dots, p_n$. Una formula proposizionale ϕ è costruita con $\wedge, \vee, \neg$ su queste atomiche. Si considerano mondi possibili W come assegnazioni di verità: ogni $w \in W$ può essere identificato con un vettore $(v_1, \dots, v_n) \in \{0,1\}^n$ che dice quali atomi sono veri in w . Una valutazione classica $\pi : W \times \{p_1, \dots, p_n\} \rightarrow \{0,1\}$ assegna ad ogni coppia (w, p_i) un valore, e si estende composizionalmente a tutte le formule proposizionali. Una distribuzione di probabilità $\mu : W \rightarrow [0,1]$ con $\sum_{w \in W} \mu(w) = 1$ permette di definire la probabilità di ϕ come:

$$\Pr(\phi) = \sum_{w : (W, \pi, w) \models \phi} \mu(w).$$

Fagin, Halpern e Megiddo estendono il linguaggio proposizionale introducendo termini probabilistici. Costruiscono un linguaggio L_P in cui compaiono espressioni come $\Pr(\phi) \geq r$, e più in generale combinazioni lineari come:

$$a_1 \Pr(\phi_1) + \dots + a_k \Pr(\phi_k) \geq b,$$

dove a_i, b sono numeri razionali. Una formula di L_P è una combinazione booleana di tali vincoli. La semantica è: un modello è una coppia (W, μ, π) come sopra; si valuta ogni vincolo sostituendo a $\Pr(\phi_i)$ il valore definito e verificando la disuguaglianza lineare.

Il problema logico fondamentale è la soddisfacibilità: data una formula Φ di L_P , esiste un modello (W, μ, π) che la soddisfa?

Si può ridurre questa domanda a un problema di programmazione lineare. Ci sono al più 2^n assegnazioni di verità distinte alle atomiche; indichiamole $v_1, \dots, v_m$, dove $m = 2^n$. Ogni mondo w è equivalente a una di queste assegnazioni; non perdiamo generalità considerando $W = \{v_1, \dots, v_m\}$. Definiamo variabili reali $x_j = \mu(v_j)$ per $j=1, \dots, m$. I vincoli di base sono:

$$x_j \geq 0 \text{ per ogni } j,$$

$$\sum_{j=1}^m x_j = 1.$$

Per ogni formula proposizionale ϕ , l’insieme $S_\phi = \{j \mid v_j \models \phi\}$ è noto; dunque $\Pr(\phi) = \sum_{j \in S_\phi} x_j$. Ogni vincolo lineare della forma $a_1 \Pr(\phi_1) + \dots + a_k \Pr(\phi_k) \geq b$ si traduce in:

$$a_1 \sum_{j \in S\{\varphi_1\}} x_j + \dots + a_k \sum_{j \in S\{\varphi_k\}} x_j \geq b,$$

cioè in una disuguaglianza lineare sui x_j . L'insieme di tutte le disuguaglianze derivanti da Φ , insieme ai vincoli di normalizzazione, definisce un sistema $S(\Phi)$ di inequazioni lineari in m variabili. La formula Φ è soddisfacibile se e solo se il sistema $S(\Phi)$ ha una soluzione reale con $x_j \geq 0$ e $\sum x_j = 1$. Dunque il problema logico è equivalente a decidere la fattibilità di un sistema lineare.

Questo mostra che il “semantico” di L_P è isomorfo alla geometria convessa di un politopo in $\mathbb{R}^m$: l'insieme delle distribuzioni di probabilità sulle assegnazioni atomiche che soddisfano i vincoli. Questo è perfettamente in linea con la prospettiva Scott–Krauss: misure logiche nel caso proposizionale sono punti in un simplex finito, e vincoli logici diventano vincoli lineari.

Nilsson affronta un problema affine ma formulato al contrario: dato un insieme di formule $\varphi_1, \dots, \varphi_k$ ciascuna con un intervallo di probabilità $[l_i, u_i]$, si chiede se esista una distribuzione di probabilità p sulle assegnazioni $v_1, \dots, v_m$ tale che, per ogni i :

$$l_i \leq \Pr(\varphi_i) \leq u_i,$$

$$\text{dove } \Pr(\varphi_i) = \sum_{j \in S\{\varphi_i\}} p(v_j).$$

Ancora una volta, si traduce tutto in disuguaglianze lineari:

$$p_j \geq 0 \text{ per tutti } j,$$

$$\sum_j p_j = 1,$$

$$l_i \leq \sum_{j \in S\{\varphi_i\}} p_j \leq u_i \text{ per ogni } i.$$

Le soluzioni p formano un politopo convesso $K \subseteq \mathbb{R}^m$. L'inferenza probabilistica si riduce a problemi di ottimizzazione lineare sui vertici di K : per trovare il minimo possibile di $\Pr(\psi)$ compatibile con i vincoli, si risolve il problema:

minimizzare $\sum_{j \in S\psi} p_j$ soggetto ai vincoli su K ;

per il massimo, un analogo problema di massimizzazione. Il fatto che K sia convesso garantisce che i valori estremi siano realizzati ai vertici, quindi è sufficiente considerare soluzioni base.

Passando alle logiche del primo ordine con quantificatori probabilistici, Keisler propone di estendere la sintassi con un quantificatore $(P x \geq r) \varphi(x)$, interpretabile come: “la misura dell'insieme di x tali che $\varphi(x)$ è vera è almeno r ”. A differenza della situazione astratta di Gaifman, qui la misura è sul dominio del modello, non sullo spazio dei modelli.

Una struttura per questa logica è una tripla $M = (D, I, \mu_D)$, dove (D, I) è una struttura per L e μ_D è una misura di probabilità definita su un anello (o σ -algebra) di sottoinsiemi di D , che includa tutti gli insiemi definibili del tipo $\{a \in D \mid M \models \varphi(a)\}$ per formule $\varphi(x)$ con parametri. La semantica è:

$$M \models (P x \geq r) \varphi(x) \text{ se e solo se } \mu_D(\{a \in D \mid M \models \varphi(a)\}) \geq r.$$

Keisler formula un sistema di assiomi che estende la logica del primo ordine ordinaria, aggiungendo schemi come:

1. $(P\ x \geq 0) \varphi(x)$ è sempre un teorema;
2. per $s < r$, $(P\ x \geq r) \varphi(x) \rightarrow (P\ x \geq s) \varphi(x)$ è un teorema;
3. un assioma di additività: se $(P\ x \geq r) \varphi(x)$, $(P\ x \geq s) \psi(x)$, e $(P\ x \geq 1) (\neg\varphi(x) \vee \neg\psi(x))$ sono veri, allora $(P\ x \geq \min(1, r+s)) (\varphi(x) \vee \psi(x))$ è un teorema;
4. monotonicità: se $\models \forall x (\varphi(x) \rightarrow \psi(x))$, allora $(P\ x \geq r) \varphi(x) \rightarrow (P\ x \geq r) \psi(x)$ è un teorema.

Il teorema di completezza di Keisler afferma che questo sistema è completo rispetto a una classe specifica di strutture misurate: le strutture di Loeb ottenute da modelli non standard iperfiniti. Intuitivamente, si usa la non standard analysis per costruire domini “molto grandi ma finiti” in senso interno, dotati della misura uniforme di conteggio interna, e poi si applica la trasformazione di Loeb per ottenere una misura standard σ -additiva.

Per spiegare la misura di Loeb con più precisione: si consideri un universo non standard in cui, per ogni insieme standard X , esiste un'estensione interna X . Sia D un insieme interno iperfinito, cioè esiste un iperintero $N^* \in \mathbb{N}$ e una biiezione interna $f: D \rightarrow \{1, \dots, N\}$. Definiamo per ogni sottoinsieme interno $A \subseteq D$ la misura interna

$$\mu_{\text{int}}(A) = \text{card}_{\text{int}}(A) / \text{card}_{\text{int}}(D^*) \in {}^*[0,1],$$

dove card_{int} è la cardinalità interna (un iperintero). μ_{int} è finitamente additiva sugli insiemi interni. La trasformazione di Loeb costruisce una misura esterna $\mu_L: \Sigma_L \rightarrow [0,1]$ su una σ -algebra Σ_L che contiene tutti gli insiemi interni, tale che, per ogni A interno, $\mu_L(A) = \text{st}(\mu_{\text{int}}(A))$, la parte standard di $\mu_{\text{int}}(A)$. La costruzione usa il procedimento di Carathéodory: definisce una outer measure su tutti i sottoinsiemi di D^* prendendo infimi di somme di $\text{st}(\mu_{\text{int}})$ su ricoprimenti interni, e poi restringe agli insiemi misurabili alla Carathéodory. Il risultato è una misura di probabilità classica su (D^*, Σ_L) .

Quando $M^* = (D, I, \mu_L)$ è una struttura di Loeb, i quantificatori $(P\ x \geq r) \varphi(x)$ si interpretano rispetto a μ_L . La completezza di Keisler consiste nel mostrare che, per ogni insieme di formule Σ consistente con gli assiomi, esiste una struttura di Loeb M^* tale che tutte le formule di Σ sono vere in M^* . Andando al dettaglio del metodo: si costruisce una sequenza di strutture finite $M_i = (D_i, I_i)$ che approssimano sempre meglio Σ , in modo che i vincoli probabilistici su misure di insiemi definibili siano approssimati dalle frequenze nei domini D_i . Si forma l'ultraprodotto $M^\# = \prod_i M_i / U$ con un ultrafiltro non principale U ; il dominio $D^\#$ è iperfinito, e abbiamo una misura interna definita dalle frequenze relative. La misura di Loeb μ_L su $D^\#$ viene definita come sopra. Usando il lemma di Łoś, si verifica che le formule di Σ sono soddisfatte in $M^\#$, e i vincoli probabilistici sono rispettati dalla misura di Loeb. Di qui la completezza.

Accanto a questo, esiste un altro teorema fondamentale di Gaifman, che riguarda la possibilità di estendere insiemi consistenti di enunciati probabilistici a misure complete su spazi di modelli. Una formulazione tipica è: sia L un linguaggio del primo ordine numerabile, e sia E un insieme di formule chiuse di L , consistente (non deriva una contraddizione in logica classica). Allora esiste una misura di probabilità μ su $S(L)$, lo spazio di Stone delle teorie complete di L , tale che $\mu([\varphi]) = 1$ per ogni $\varphi \in E$, dove $[\varphi] = \{ U \in S(L) \mid \varphi \in U \}$. In modo equivalente, E può essere “realizzato con probabilità 1” da una misura su teorie complete. Il proof idea è quello della costruzione di una misura prodotto su un'algebra booleana libera e della sua spinta in avanti sull'algebra di Lindenbaum–Tarski, usando lemmi di indipendenza tipo Rasiowa–Sikorski: si costruisce una famiglia di eventi indipendenti che generano $LT(L)$ e si definisce una misura prodotto; l'estensione a E usa il fatto che E è consistente, quindi ogni vincolo $\varphi \in E$ può essere visto come un evento di misura 1.

Hoover, infine, lavora sulle misure su strutture di L con dominio numerabile, e ne studia due proprietà: definibilità ed exchangeability (scambiabilità). Una misura μ su strutture di L con dominio $\mathbb{N}$ è scambiabile se per ogni permutazione $\pi : \mathbb{N} \rightarrow \mathbb{N}$ e ogni insieme misurabile A di strutture vale $\mu(A) = \mu(A^\pi)$, dove $A^\pi = \{ M^\pi \mid M \in A \}$ e M^π è la struttura ottenuta da M permutando il dominio con π . La definibilità, a grandi linee, significa che la misura μ è “prodotta” da proprietà quantificazionali che possono essere approssimate da combinazioni di fatti atomici: per ogni formula $\varphi(x)$, e ogni $\varepsilon > 0$, esiste una formula quantificatore-free $\psi(x)$ tale che, per μ -quasi tutte le strutture M , la percentuale di elementi $a \in D^M$ che soddisfano $\varphi(a) \leftrightarrow \psi(a)$ è almeno $1-\varepsilon$. In simboli, per quasi tutti M :

$$\mu_M(\{ a \mid M \models \varphi(a) \leftrightarrow \psi(a) \}) \geq 1-\varepsilon,$$

dove μ_M è una misura naturale sul dominio (conteggio normalizzato su insiemi finiti, o misura di Loeb in casi non standard). Hoover mostra che le misure scambiabili e definibili hanno rappresentazioni come miscele di misure prodotto, analoghe alla rappresentazione di de Finetti e all'estensione Aldous–Hoover: esistono parametri latenti θ in qualche spazio Θ e una misura a priori ν su Θ tali che

$$\mu = \int \mu_\theta \, d\nu(\theta),$$

dove μ_θ è una misura su strutture di L in cui i fatti atomici (come $R(a_1, \dots, a_n)$) sono indipendenti condizionatamente a θ , con distribuzioni determinate da θ .

Questa linea di rappresentazioni porta direttamente alle logiche infinitarie e alle leggi zero–uno. Nella logica $L_{\{\omega, 1\}}$, si consentono congiunzioni e disgiunzioni numerabili, mantenendo quantificatori finiti. Per strutture casuali come il grafo $G(n, p)$ con $0 < p < 1$ fisso, si dimostra che per il frammento FO (e per alcune estensioni a $L_{\{\omega, 1\}, P}$, dove P indica quantificatori di probabilità) vale una legge zero–uno: per ogni formula φ in quel frammento, il limite

$$\lim_{n \rightarrow \infty} \Pr(G(n, p) \models \varphi)$$

esiste e vale 0 o 1. La dimostrazione tipica usa giochi di Ehrenfeucht–Fraïssé: per un dato numero di mosse k , si considera un gioco fra due giocatori su due grafi, e si dimostra che con alta probabilità, per n grande, tutti i grafi $G(n, p)$ “sembrano uguali” rispetto a giochi di lunghezza k . Questo implica che tutte le formule con profondità quantificazionale $\leq k$ hanno valore di verità asintotico determinato. Passando a $k \rightarrow \infty$, si ottiene la completezza della teoria quasi-sicura. Quando si aggiungono quantificatori probabilistici, la situazione è più sottile: si possono esprimere proprietà di densità, ma per certe combinazioni (ad esempio quantificatori che misurano la densità di copie finite di un sottografo) esistono ancora leggi zero–uno raffinate.

Un tema distinto ma collegato è quello delle probabilità infinitesimali e delle funzioni di Popper. Una funzione di Popper P è definita su coppie (A, B) di eventi (sottoinsiemi di un universo Ω) con B non vuoto, e produce un valore $P(A|B)$ in un campo K esteso, spesso non standard, in modo che per ogni B fisso, $A \mapsto P(A|B)$ sia una probabilità (finitamente additiva) e che si abbia un assioma di catena tipo:

$$P(A \wedge C \mid B) = P(A|B) P(C \mid A \wedge B),$$

quando $A \wedge B$ non è impossibile. Se P viene da una probabilità incondizionata P_0 via $P(A|B) = P_0(A \wedge B)/P_0(B)$, allora le funzioni di Popper non aggiungono nulla. Ma se si permette che $P_0(B)$ sia infinitesimale, allora $P(A|B)$ può essere definito anche quando $P_0(B) = 0$ in senso standard.

Modelli non standard della probabilità, come quelli costruiti con numeri iperreali, permettono di definire P_0 con valori in $^*[0,1]$ e di trattare gli eventi B deboli (di misura infinitesimale ma non zero) con condizionali ben definiti. Le logiche condizionali probabilistiche usano funzioni di Popper per interpretare enunciati del tipo “se ϕ allora ψ ” come affermazioni su $P(\psi|\phi)$ che è “quasi 1” (1 meno un infinitesimo). In questo modo, “verità condizionale” assume una dimensione probabilistica fine.

Accanto a questi aspetti tecnici, ci sono premesse set-teoriche profonde. L’assioma della scelta pieno (AC) implica l’esistenza di sottoinsiemi non misurabili di $[0,1]$ rispetto alla misura di Lebesgue, a meno di ipotesi aggiuntive forti. Solovay ha mostrato che se si assume l’esistenza di un cardinale inaccessibile, si può costruire un modello di $ZF + DC$ in cui ogni sottoinsieme di $\mathbb{R}$ è Lebesgue misurabile. In tale modello, la probabilità di Lebesgue è “totale”: ogni sottoinsieme ha una misura. Questo ha un forte fascino per le fondazioni probabilistiche, ma è incompatibile con AC completo. Dal punto di vista della logica probabilistica, questo significa che un ideale “universo misura-teorico” in cui ogni proprietà definibile di $\mathbb{R}$ è misurabile richiede ipotesi set-teoriche forti; in ZFC standard, bisogna accontentarsi della misurabilità per classi meno generali (come gli insiemi boreliani o proiettivi sotto ipotesi di determinacy). Questa tensione tra ambizioni di completezza probabilistica e limiti set-teorici è un fatto strutturale: se si vuole una probabilità definita su “tutte” le definizioni logiche possibili, si entra in una zona dove le ipotesi di grandi cardinali diventano rilevanti.

Infine, il collegamento con la teoria dell’apprendimento e la complessità: la semantica PAC (Probably Approximately Correct) è una forma di logica probabilistica operativa, in cui si chiede che una regola (o formula) sia “quasi sempre vera” rispetto a distribuzioni a priori sconosciute, con vincoli sulle risorse di campionamento. Una classe di concetti C (sottoinsiemi di X) è PAC-learnable se esiste un algoritmo che, per ogni $c \in C$, ogni distribuzione D su X , e ogni $\varepsilon, \delta > 0$, con $m = \text{poly}(1/\varepsilon, 1/\delta)$ esempi i.i.d. da D etichettati secondo c produce un’ipotesi h tale che con probabilità $\geq 1-\delta$ si ha $\Pr_{x \sim D}[h(x) \neq c(x)] \leq \varepsilon$. La dimensione VC di C è il massimo k tale che esiste un insieme $S \subseteq X$ di cardinalità k che viene “frammentato” completamente da C : per ogni etichettatura 0/1 di S , esiste un concetto in C che produce esattamente quella etichetta su S . Il teorema fondamentale (Blumer–Ehrenfeucht–Haussler–Warmuth) è che C è PAC-learnable se e solo se $\text{VCdim}(C)$ è finita. Applicato a classi definibili logicamente, questo significa che frammenti di una logica (per esempio, FO con profondità quantificazionale e arietà limitata) producono classi di concetti di VC-dimension finita, mentre FO completo produce tipicamente VC-dimension infinita. In termini logici, questo vuol dire che la potenza espressiva di una logica influisce direttamente sulla capacità di generalizzare da dati finiti, vista probabilisticamente.

La fondazione categoriale chiude questo Volume I: il funtore di Giry G che manda uno spazio misurabile X nello spazio $G(X)$ delle misure di probabilità su X , con unità $\eta_X : X \rightarrow G(X)$ che manda x in δ_x (la delta di Dirac) e moltiplicazione $\mu_X : G(G(X)) \rightarrow G(X)$ che integra misure di misure, definisce una monade. Le categorie di Markov sono categorie in cui i morfismi rappresentano trasformazioni probabilistiche: un morfismo $f : X \rightarrow Y$ è qualcosa che, a ogni $x \in X$, associa una misura $f(x, \cdot)$ su Y . La composizione di morfismi è l’integrazione delle misure. Molte costruzioni che abbiamo visto – miscele di prodotti, aggiornamento bayesiano, scambiabilità come invarianza sotto permutazioni – possono essere riformulate come proprietà di morfismi in una Markov category: l’aggiornamento bayesiano è l’applicazione di μ_X , la miscela è l’applicazione di μ su $G(G(X))$, la simmetria di de Finetti è una condizione di invarianti sotto l’azione del gruppo delle permutazioni, ecc. In questa ottica, la logica probabilistica può essere vista come l’insieme degli enunciati che rimangono invarianti sotto certe classi di morfismi probabilistici: un enunciato è “logico” se è preservato da tutti i functor che rispettano la struttura monoidale e la monade di Giry.

Consideriamo L un linguaggio del primo ordine numerabile. L'insieme delle formule chiuse (senza variabili libere) è numerabile. Indichiamo con $\text{Sent}(L)$ l'insieme delle formule chiuse. Se φ è una formula chiusa, scriviamo $[\varphi]$ per l'insieme dei modelli M di L tali che $M \models \varphi$.

Definizione. Sia Q un insieme rappresentativo di strutture per L (a isomorfismo trascurato). Per ogni formula chiusa φ , definiamo:

$$[\varphi] = \{ M \in Q \mid M \models \varphi \}.$$

Lasciamo che $\mathcal{F}_0$ sia la famiglia di tutti gli insiemi $[\varphi]$. Lasciamo $\mathcal{F}$ la più piccola algebra booleana di sottoinsiemi di Q che contiene $\mathcal{F}_0$ (cioè chiusa sotto unioni finite, intersezioni finite e complementi). Lasciamo Σ_L la σ -algebra generata da $\mathcal{F}$.

Definizione (misura logica semantica). Una misura logica semantica è una misura di probabilità $\mu : \Sigma_L \rightarrow [0,1]$ tale che, per tutte le formule φ, ψ :

1. $\mu([\top]) = 1, \mu([\perp]) = 0$;
2. $\mu([\varphi \vee \psi]) = \mu([\varphi]) + \mu([\psi]) - \mu([\varphi \wedge \psi])$.

Osserviamo che (2) è solo la scrittura del fatto che $[\varphi \vee \psi] = [\varphi] \cup [\psi]$ e $[\varphi \wedge \psi] = [\varphi] \cap [\psi]$, combinato con la formula di additività finita per le misure: $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.

Il problema non è l'esistenza di misure su Σ_L in astratto (perché qualunque algebra booleana numerabile ammette misure, per esempio si può mettere una misura di conteggio smorzata su un numero finito di atomi e poi estendere), il problema è l'interazione con i quantificatori, che vogliamo esprimere tramite limiti sopra nuove costanti.

Introduciamo l'algebra di Lindenbaum–Tarski.

Definizione. Consideriamo la relazione di equivalenza logica su $\text{Sent}(L)$: $\varphi \equiv \psi$ se e soltanto se $\models \forall (\varphi \leftrightarrow \psi)$, cioè se per ogni modello M , $M \models \varphi \leftrightarrow \psi$. Denotiamo con $[\varphi]_{LT}$ la classe di equivalenza di φ . L'insieme $LT(L) = \text{Sent}(L) / \equiv$, con operazioni:

$$[\varphi]_{LT} \wedge [\psi]_{LT} = [\varphi \wedge \psi]_{LT}, [\varphi]_{LT} \vee [\psi]_{LT} = [\varphi \vee \psi]_{LT}, \neg[\varphi]_{LT} = [\neg\varphi]_{LT},$$

è un'algebra booleana. Gli elementi di $LT(L)$ sono le formule modulo equivalenza logica.

Definizione (misura logica sintattica). Una funzione $\mu_L : LT(L) \rightarrow [0,1]$ è una misura logica se:

1. $\mu_L([\top]_{LT}) = 1, \mu_L([\perp]_{LT}) = 0$;
2. se $[\varphi]_{LT} \wedge [\psi]_{LT} = [\perp]_{LT}$, allora $\mu_L([\varphi]_{LT} \vee [\psi]_{LT}) = \mu_L([\varphi]_{LT}) + \mu_L([\psi]_{LT})$;
3. se $\{ [\varphi_i]_{LT} \}$ è una famiglia numerabile di elementi mutuamente disgiunti (cioè $\varphi_i \wedge \varphi_j$ è contraddittoria per $i \neq j$), allora

$$\mu_L(\bigvee_i [\varphi_i]_{LT}) = \sum_i \mu_L([\varphi_i]_{LT}).$$

Notare che questa è esattamente la definizione di misura di probabilità su una σ -algebra, tradotta nel linguaggio della algebra booleana $LT(L)$.

Teorema (rappresentazione di Scott–Krauss). Sia $\mu_L : LT(L) \rightarrow [0,1]$ una misura logica. Allora esiste uno spazio di probabilità (Ω, Σ, P) e, per ogni formula chiusa φ , una funzione misurabile $f_\varphi : \Omega \rightarrow \{0,1\}$ tali che, per ogni φ :

$$\mu_L([\varphi] \{LT\}) = P(\{ \omega \in \Omega \mid f_\varphi(\omega) = 1 \}).$$

Inoltre, la σ -algebra generata dagli insiemi $\{ \omega \mid f_\varphi(\omega) = 1 \}$ è densa modulo insiemi di P-misura zero, e la rappresentazione è unica a meno di isomorfismi di algebre di misura.

Idea della dimostrazione. Si considera l'ideale degli elementi di misura zero:

$$I = \{ a \in LT(L) \mid \mu_L(a) = 0 \}.$$

Si quozienta $LT(L)$ per I , ottenendo una algebra $B = LT(L)/I$. La misura μ_L induce una misura μ_B su B ponendo $\mu_B(a/I) = \mu_L(a)$. È facile verificare che μ_B è ben definita e σ -additiva, e che (B, μ_B) è un'algebra di misura. Esiste un teorema generale: ogni algebra di misura completa e atomless può essere rappresentata come algebra di insiemi misurabili modulo nulli di uno spazio di probabilità standard (questo è un risultato classico della teoria delle algebre di misura; una forma è l'esistenza di uno spazio di probabilità standard le cui classi modulo nulli sono isomorfe a B). Quindi esiste uno spazio di probabilità (Ω, Σ, P) e un isomorfismo di algebre di misura $\theta : B \rightarrow \Sigma \text{mod } 0$, dove $\Sigma \text{mod } 0$ è l'insieme delle classi modulo insiemi di misura zero degli elementi di Σ . Per ogni formula φ , $[\varphi] \{LT\}$ appartiene a $LT(L)$; la sua immagine in B è $[\varphi] \{LT\} / I$. La sua immagine in $\Sigma \text{mod } 0$ è la classe di qualche $A_\varphi \in \Sigma$ modulo nulli. Definiamo f_φ come la funzione caratteristica di A_φ . Allora

$$P(\{ \omega \mid f_\varphi(\omega) = 1 \}) = P(A_\varphi) = \mu_B([\varphi] \{LT\} / I) = \mu_L([\varphi] \{LT\}).$$

L'unicità a meno di isomorfismi segue dall'unicità (a meno di isomorfismo) dei modelli delle algebre di misura date.

Questo teorema dice in sostanza: parlare di “ $\mu_L([\varphi])$ ” come probabilità logica di φ è equivalentemente parlare della probabilità P che una variabile casuale booleana f_φ valga 1 in uno spazio di probabilità classico.

Da qui, la possibilità di costruire misure logiche “semanticamente” (sullo spazio dei modelli) e poi rappresentarle “analiticamente”, e viceversa.

Adesso entriamo nel dettaglio della costruzione non standard e della misura di Loeb, perché è il pezzo tecnico centrale per la logica con quantificatori di probabilità (Keisler) e per molte semantiche probabilistiche su domini infiniti.

L'idea di base della non standard analysis è ottenere un'estensione ${}^*\mathbb{R}$ di $\mathbb{R}$ in cui esistono infinitesimi (numeri > 0 ma minori di $1/n$ per ogni n standard) ed elementi infiniti (maggiore di ogni $n \in \mathbb{N}$). Uno dei modi classici di costruire ${}^*\mathbb{R}$ è con un ultraprodotto: si prende una famiglia $(\mathbb{R})_i$ indicizzata da $\mathbb{N}$, si forma il prodotto $\prod_i \mathbb{R}$, e si quozienta rispetto a un ultrafiltro non principale U su $\mathbb{N}$. Due sequenze (x_i) e (y_i) sono equivalenti se $\{ i \mid x_i = y_i \} \in U$. L'insieme delle classi di equivalenza è ${}^*\mathbb{R}$. Un numero standard $a \in \mathbb{R}$ viene identificato con la classe della sequenza costante $(a, a, a, \dots)$. Un numero $x \in {}^*\mathbb{R}$ è infinito se per ogni n standard, $|x| > n$; è infinitesimo se per ogni n standard, $|x| < 1/n$. È un fatto che ogni x finito in ${}^*\mathbb{R}$ (cioè non infinito) è infinitamente vicino a un unico numero reale $\text{st}(x)$: questo è la funzione “standard part”.

Ora prendiamo una sequenza di strutture finite $M_i = (D_i, I_i)$ per $i \in \mathbb{N}$, con $|D_i| = N_i$. Costruiamo l'ultraprodotto $M^* = \prod_i M_i / U$. Il dominio D^* è l'ultraprodotto dei D_i , che è un insieme interno in V (l'universo non standard) che può essere messo in biiezione interna con un segmento $\{1, \dots, N\}$ di $\mathbb{N}$, per qualche iperintero N infinito. Definiamo, per ogni sottoinsieme interno $A \subseteq D^*$, la cardinalità interna $\text{card_int}(A)$ come l'ultraprodotto delle cardinalità $|A_i|$ dei corrispondenti $A_i \subseteq D_i$. Definiamo la misura interna:

$$\mu_{\text{int}}(A) = \text{card_int}(A) / \text{card_int}(D^*),$$

che è un elemento di $[0,1]$. μ_{int} è finitamente additiva: se A, B sono interni e disgiunti, $\mu_{\text{int}}(A \cup B) = \mu_{\text{int}}(A) + \mu_{\text{int}}(B)$, perché $\text{card_int}(A \cup B) = \text{card_int}(A) + \text{card_int}(B)$.

Il passaggio alla misura di Loeb consiste nel definire una misura esterna μ_L con valori in $[0,1]$, σ -additiva, che estende la parte standard di μ_{int} sugli insiemi interni.

Definizione (outer measure di Loeb). Per ogni sottoinsieme arbitrario $E \subseteq D^*$, definiamo:

$$\mu^*(E) = \inf \{ \sum_n \text{st}(\mu_{\text{int}}(A_n)) \mid E \subseteq \bigcup_n A_n, A_n \text{ interni} \}.$$

Questo è un'outer measure: $\mu^*(\emptyset) = 0$, $\mu^*(E) \leq \mu^*(F)$ se $E \subseteq F$, e $\mu^*(\bigcup_i E_i) \leq \sum_i \mu^*(E_i)$. Per definizione, per ogni insieme interno A , $\mu^*(A) \leq \text{st}(\mu_{\text{int}}(A))$, e in realtà si può dimostrare che $\mu^*(A) = \text{st}(\mu_{\text{int}}(A))$ (perché si può prendere la copertura con un unico insieme A).

Definizione (insiemi misurabili alla Carathéodory). Un insieme $B \subseteq D^*$ è detto μ^* -misurabile se, per ogni $E \subseteq D^*$, si ha:

$$\mu^*(E) = \mu^*(E \cap B) + \mu^*(E \setminus B).$$

L'insieme Σ_L degli insiemi misurabili forma una σ -algebra, e la restrizione $\mu_L = \mu^*|_{\Sigma_L}$ è una misura di probabilità completa. Per insiemi interni A , A è misurabile e $\mu_L(A) = \text{st}(\mu_{\text{int}}(A))$.

Questo è il contenuto del teorema di Loeb. Il teorema formale suona così:

Teorema (Loeb). Sia D^* un insieme interno iperfinito in un universo non standard, e μ_{int} una misura interna finitamente additiva su sottoinsiemi interni di D^* con $\mu_{\text{int}}(D^*) = 1$. Allora esiste una σ -algebra Σ_L e una misura σ -additiva $\mu_L : \Sigma_L \rightarrow [0,1]$ tale che:

1. ogni sottoinsieme interno A di D^* è in Σ_L ;
2. per ogni A interno, $\mu_L(A) = \text{st}(\mu_{\text{int}}(A))$;
3. μ_L è completa.

Questa costruzione fa da ponte perfetto con Keisler: i quantificatori probabilistici $(P x \geq r) \varphi(x)$ vengono interpretati in strutture $M^* = (D, I, \mu_L)$ in cui D^* e I^* sono non standard (ultraprodotti di strutture finite) e μ_L è la misura di Loeb.

Teorema (Keisler, completezza). Sia L_P una logica del primo ordine estesa con quantificatori di probabilità $(P x \geq r) \varphi(x)$, con r razionale in $[0,1]$. Sia K la classe di tutte le strutture di Loeb $M^* = (D, I, \mu_L)$ che interpretano L_P come sopra. Allora il sistema assiomatico di Keisler (assiomi FO classici, più gli schemi 1–4 sui quantificatori P) è completo rispetto a K : per ogni formula χ di L_P , se χ è vera in tutte le strutture di K , allora χ è dimostrabile nel sistema; se χ non è dimostrabile, esiste una $M^* \in K$ tale che $M^* \not\models \chi$.

Idea della dimostrazione. Se χ non è dimostrabile, l'insieme $\Sigma = \{\neg\chi\} \cup \text{Th}$ è consistente, dove Th contiene tutti gli assiomi e tutte le loro conseguenze logiche. Si estende Σ a una teoria massimale T. Si costruisce una catena di strutture finite M_i che approssimano T: su ciascuna M_i , le formule con quantificatori di probabilità sono interpretate attraverso frequenze di elementi nel dominio; cioè si definisce una misura $\mu_i(A) = |A| / |D_i|$ per sottoinsiemi A definibili. Si cura di scegliere M_i in modo che, per ogni formula di T, il valore approssimato dalle frequenze sia vicino ai vincoli imposti dai quantificatori ($P x \geq r$). Il lemma di compattezza del primo ordine assicura che tali approssimazioni finite esistono per insiemi finiti di formule. Passando a un ultraprodotto $M^\# = \prod_i M_i / U$, otteniamo una struttura non standard $M^\#$ con dominio $D^\#$, misura interna μ_{int} costruita da μ_i , e interpretazione $I^\#$. La misura di Loeb μ_L su $D^\#$ definisce una struttura $M^* = (D^\#, I^\#, \mu_L)$. Si verifica che M^* soddisfa T e quindi soddisfa $\neg\chi$; dunque χ non è valida in K.

Questa struttura è elegante perché dimostra la robustezza della logica con quantificatori probabilistici: malgrado la presenza di misure, limiti e numeri reali, il sistema rimane modellabile in un quadro altamente controllato (strutture iperfinite e misure di Loeb).

Passiamo ora alla rappresentazione scambiabile alla Aldous–Hoover e al ruolo di Hoover per misure su strutture.

Consideriamo dapprima il caso delle sequenze. Una famiglia di variabili casuali $\{X_i\}_{i \in \mathbb{N}}$ a valori in un insieme S è scambiabile se, per ogni permutazione finita π di $\mathbb{N}$, la distribuzione congiunta di (X_i) è uguale a quella di $(X_{\pi(i)})$. Il teorema di de Finetti dice che, se $\{X_i\}$ è scambiabile e S è discreto, allora esiste una misura a priori π su $\Delta(S)$ (lo spazio delle distribuzioni su S) tale che la legge di $\{X_i\}$ è uguale a una miscela di processi i.i.d:

$$\Pr(X_1 = x_1, \dots, X_n = x_n) = \int \prod_{i=1}^n \mu(x_i) d\pi(\mu).$$

Nel caso di $S = \{0,1\}$, questo si riduce a una misura π su $[0,1]$, e spesso si scrive:

$$\Pr(X_1 = x_1, \dots, X_n = x_n) = \int 0^{s_1} 1^{1-s_1} \theta^{s_2} (1-\theta)^{n-s_2} d\pi(\theta),$$

dove s è il numero di i con $x_i = 1$.

Spostiamoci ora su array bidimensionali $\{X_{ij}\}_{i,j \in \mathbb{N}}$, con X_{ij} a valori in S. Una distribuzione su tali array è detta scambiabile se per ogni permutazione finita π di $\mathbb{N}$,

$$(X_{ij})_{i,j \in \mathbb{N}} \equiv (X_{\pi(i) \pi(j)})_{i,j \in \mathbb{N}}$$

in distribuzione. Il teorema di Aldous–Hoover generalizza de Finetti: esistono variabili casuali indipendenti U, $\{U_i\}$, $\{V_j\}$, $\{W_{ij}\}$, tutte uniformi su $[0,1]$, e una funzione misurabile $f : [0,1]^4 \rightarrow S$ tale che

$$X_{ij} = f(U, U_i, V_j, W_{ij})$$

in distribuzione. Di solito, nel caso di array simmetrici, si richiede anche $X_{ij} = X_{ji}$, che impone simmetrie su f.

Questo ha una lettura logica se pensiamo a X_{ij} come al valore di verità di un predicato binario $R(i,j)$. Un modello random scambiabile per un linguaggio con un solo simbolo binario R può essere visto come generato da parametri latenti U (globale) e U_i (per ogni elemento del dominio), e da rumore indipendente W_{ij} , attraverso una funzione f che determina se $R(i,j)$ sia vero o falso.

Condizionatamente ai parametri U, U_i , le variabili $R(i,j)$ sono indipendenti. La misura su strutture è dunque una miscela (integrale su U, U_i) di misure prodotto condizionate.

Hoover generalizza questa idea a linguaggi arbitrari. Per semplicità, pensiamo a un linguaggio finito con relazioni di arità $\leq k$. Una struttura M con dominio $\mathbb{N}$ è determinata dai valori $R(a_1, \dots, a_n)$ per ogni simbolo R di arità n e ogni n -upla di elementi del dominio. Una misura scambiabile su tali strutture è una misura su $\{0,1\}^{\mathbb{I}}$, dove $\mathbb{I}$ è l'insieme di tutti gli indici $(R, (a_1, \dots, a_n))$, invariante sotto permutazioni di $\mathbb{N}$. Il teorema di Hoover (in stile Aldous–Hoover) afferma che esiste un sistema di variabili latenti e funzioni misurabili che generano la distribuzione dei fatti atomici a partire da rumore i.i.d. e parametri globali/di livello individuale. Tecnicamente, si usa una catena di rappresentazioni: strutture con relazioni di arità 1, poi 2, poi 3, ecc., ciascuna rappresentata da funzioni $f_1(U, U_i, W_i), f_2(U, U_i, U_j, W_{\{ij\}}), \dots$, con vincoli di compatibilità.

Un aspetto sottile è la definibilità: affinché questa struttura rimanga nel dominio del linguaggio L , si richiede che le funzioni f siano tali che le proprietà probabilistiche delle formule di L siano determinate da combinazioni dei parametri latenti in modi esprimibili o almeno approssimabili in L . Da qui l'idea di misura definibile: una misura μ su strutture di L è definibile se, informamente, per ogni formula $\phi(x)$ e per ogni $\varepsilon > 0$, esiste una formula quantificatore-free $\psi(x)$ tali che l'insieme degli M in cui la differenza tra $\{a \mid \phi(a)\}$ e $\{a \mid \psi(a)\}$ ha “peso” almeno ε è di μ -misura piccola. Formalmente, qualcosa come:

$$\mu(\{M \mid \mu_M(\{a \mid M \models \phi(a) \leftrightarrow \psi(a)\}) \geq 1 - \varepsilon\}) \geq 1 - \varepsilon.$$

Le rappresentazioni di Hoover mostrano che misure scambiabili e definibili hanno la struttura di miscele di modelli indipendenti condizionati a parametri: un parallelo raffinatissimo di de Finetti nel mondo dei modelli di L .

Se adesso guardiamo le logiche infinitarie con probabilità, possiamo pensare alla logica $L_{\{\omega, 1\}}$ arricchita con quantificatori di probabilità, spesso denotata $L_{\{\omega, 1, P\}}$. In $L_{\{\omega, 1\}}$, ammesso un vocabolario L , è permesso formare congiunzioni $\bigwedge_{i \in I} \phi_i$ e disgiunzioni $\bigvee_{i \in I} \phi_i$ per insiemi I numerabili, mentre la lunghezza dei blocchi di quantificatori rimane finita. Se aggiungiamo la possibilità di esprimere “la misura degli x che soddisfano $\phi(x)$ è almeno r ” attraverso $(P x \geq r) \phi(x)$, otteniamo un linguaggio potentissimo per descrivere proprietà globali di strutture infinite. In modelli random come $G(n, p)$ per grafi, si dimostrano spesso leggi zero–uno: per frammenti ragionevolmente controllati di $L_{\{\omega, 1, P\}}$, per ogni formula ϕ , la probabilità che $G(n, p) \models \phi$ tende a 0 o 1 quando $n \rightarrow \infty$. Le dimostrazioni estendono tecniche note per FO: si studiano giochi di Ehrenfeucht–Fraïssé rivestiti di vincoli probabilistici, e si mostra che, per n grande, quasi tutti i grafi sono indistinguibili da un “grafo limite” rispetto a quel frammento di logica. In queste versioni, la logica non parla solo di esistenza o di universalità di strutture finite, ma anche di densità di sottostrutture (per esempio la densità di triangoli, di cammini, ecc.), e le leggi zero–uno dicono che queste densità hanno limiti deterministici.

Tutto questo apre naturalmente alla questione delle probabilità infinitesimali e delle funzioni di Popper: come interpretare logicamente condizionali su eventi di probabilità standard zero, ma non logicamente impossibili? Qui si entra in un territorio dove la non standard analysis viene riciclata in modo concettuale, piuttosto che tecnico: si considera un campo esteso K (ad esempio ${}^*\mathbb{R}$) contenente valori infinitesimali; una funzione $P : \mathcal{P} \times \mathcal{P} \setminus \emptyset \rightarrow K$ nonnulla che associa a ogni coppia (A, B) con $B \neq \emptyset$ un valore $P(A|B)$ e soddisfa le proprietà di una probabilità condizionata (additività, normalizzazione $P(B|B) = 1$, catena $P(A \wedge C|B) = P(A|B)P(C|A \wedge B)$). Quando si tenta di sviluppare una logica dei condizionali $\phi > \psi$ basata su $P(\psi|\phi)$, si ottiene un sistema in cui $\phi > \psi$ significa che $P(\psi|\phi)$ è “grande” (vicino a 1), e $\phi > \neg\psi$ significa che $P(\psi|\phi)$ è “piccolo” (vicino a 0). La finezza degli infinitesimi

consente di distinguere casi che in probabilità standard verrebbero compressi in 0 o 1, e questo ha impatto sulla trattazione dei paradossi, delle credenze revisionate, e delle nozioni di “plausibilità quasi certa”.

Arriviamo quasi al bordo del Volume I con il fronte set-teorico e apprenditivo. L’esistenza di insiemi non misurabili di reali in ZFC è un fatto noto; la visione di Solovay è che, assumendo l’esistenza di un cardinale inaccessibile, si può costruire un modello in cui ogni sottoinsieme di $\mathbb{R}$ è misurabile, ha la proprietà di Baire, ecc. In tale modello “Solovay”, per ogni $A \subseteq \mathbb{R}$ esiste $\mu_{\text{Leb}}(A)$ e non ci sono controesempi tipo Vitali. Se pensiamo alla logica probabilistica su $\mathbb{R}$ come a un sistema che assegna probabilità a ogni formula che definisce un sottoinsieme di $\mathbb{R}$, avere “tutto misurabile” è una condizione ideale. Ma questo è possibile solo rinunciando ad AC e accettando forti ipotesi sui cardinali. Si delinea quindi un principio: la pretesa che ogni enunciato logico definibile su un dominio continuo abbia una probabilità definita è, in ultima analisi, una tesi set-teorica forte, non garantita in ZFC.

Sul versante PAC, ricollochiamo l’idea con un po’ più di rigore. Sia X un insieme, C una classe di sottoinsiemi di X . Ogni $c \in C$ è un concetto. Una distribuzione D su X genera esempi (x, y) con $x \sim D$ e $y = 1$ se $x \in c$, $y = 0$ altrimenti. Un algoritmo A , osservati m esempi $(x_1, y_1), \dots, (x_m, y_m)$, produce un’ipotesi h (che può appartenere o meno a C). Definiamo l’errore generalizzato di h rispetto a c e D :

$$\text{err}_D(h, c) = \Pr_{x \sim D} [h(x) \neq c(x)].$$

Definizione. C è PAC-learnable se esiste un algoritmo A e una funzione polinomiale $m(\epsilon, \delta)$ tale che, per ogni $\epsilon, \delta > 0$, per ogni $c \in C$, per ogni D , se A riceve $m \geq m(\epsilon, \delta)$ esempi i.i.d. da D etichettati da c , allora con probabilità almeno $1 - \delta$ sull’estrazione del campione si ha $\text{err}_D(h, c) \leq \epsilon$.

Teorema (Blumer–Ehrenfeucht–Haussler–Warmuth). C è PAC-learnable se e solo se $\text{VCdim}(C) < \infty$. Inoltre, è sufficiente che $m(\epsilon, \delta) = O(\text{VCdim}(C) (\log(1/\epsilon) + \log(1/\delta)) / \epsilon^2)$.

Qui $\text{VCdim}(C)$ è la dimensione VC: il massimo k tale che esiste un insieme $S \subseteq X$ di cardinalità k che viene “frammentato” da C , cioè per ogni funzione $f: S \rightarrow \{0, 1\}$ esiste $c \in C$ con $c \cap S = f^{-1}\{1\}$. Se non esiste tale k massimo, $\text{VCdim}(C) = \infty$.

Applicata alla logica, si prende X come il dominio di strutture finite, e C come una famiglia di insiemi definibili da formule $\text{FO}(x)$ in un certo frammento. Se il frammento è abbastanza semplice (ad esempio, FO con una sola quantificazione, o con profondità quantificazionale limitata, su una struttura con arietà limitata), è possibile dimostrare che $\text{VCdim}(C)$ è finito; esistono risultati in questa direzione che collegano profondità quantificazionale, numero di predicati e crescita della VC-dimension. Questo implica che quel frammento logico è “apprendibile” in senso PAC: esiste un algoritmo che, osservando esempi, identifica quasi sempre il concetto corretto definito da una formula del frammento. Viceversa, per FO pieno, è tipico che VCdim sia infinita, e quindi il frammento non sia PAC-learnable in generale. Questo lega in modo netto potere espressivo logico e complessità apprenditiva probabilistica.

La stessa filosofia si ripete nell’apprendimento online: la famiglia di esperti può essere vista come una classe C di strategie probabilistiche (ogni teoria T definisce predizioni $\Pr(y_t | \text{passato})$ per il tempo t), e ciò che conta è il “complesso dimensionale” di C (che appare nei bound di regret).

Infine, la chiusura categoriale: la monade di Giry. Per ogni spazio misurabile (X, Σ_X) , si consideri $G(X)$ l’insieme di tutte le misure di probabilità su (X, Σ_X) , con la σ -algebra generata dalle mappe v

$\mapsto v(A)$ per $A \in \Sigma X$. Per ogni funzione misurabile $f: X \rightarrow Y$, si definisce $G(f): G(X) \rightarrow G(Y)$ spedendo v in $f^*(v)$, l'immagine diretta: $f^*(v)(B) = v(f^{-1}\{B\})$. Questa struttura (G, η, μ) con

$$\eta_X: X \rightarrow G(X), x \mapsto \delta_x,$$

$$\mu_X: G(G(X)) \rightarrow G(X), \Pi \mapsto (A \mapsto \int G(X) v(A) d\Pi(v)),$$

soddisfa gli assiomi di una monade: $\mu_X \circ G(\eta_X) = \mu_X \circ \eta\{G(X)\} = \text{id}\{G(X)\}$, $\mu_X \circ G(\mu_X) = \mu_X \circ \mu\{G(X)\}$. In termini probabilistici, η_X introduce punti di vista degenerati (delta di Dirac), μ_X “appiana” misure di misure in misure semplici.

Una categoria di Markov è una categoria in cui:

- gli oggetti sono spazi “di stati” (come X);
- i morfismi $f: X \rightarrow Y$ sono kernel di Markov (f associa ad ogni x una misura $f(x, \cdot)$ su Y);
- la composizione è definita tramite integrazione: $(g \circ f)(x, B) = \int f(x, dy) g(y, B)$;
- c'è una struttura monoidale simmetrica che permette di prendere prodotti;
- ciascun oggetto ha comonoide (copy, delete) che modella duplicazione e marginalizzazione.

In questo quadro, la logica probabilistica può essere vista come la teoria di invarianti: enunciati che restano veri sotto l'azione di certi morfismi. De Finetti e Aldous–Hoover, ad esempio, sono teoremi di rappresentazione che si possono riformulare come fatti su morfismi simmetrici in una Markov category: un morfismo $X^{\mathbb{N}} \rightarrow Y$ scambiabile può essere fattorizzato come un morfismo $X^{\mathbb{N}} \rightarrow Z$ (che rappresenta il parametro latente) seguito da un morfismo $Z \rightarrow Y^{\mathbb{N}}$ che è prodotto di un morfismo elementare.

L'idea che la probabilità sia un'estensione naturale della logica trova la sua piena maturazione quando si passa dai fondamenti astratti alle strutture operative che popolano l'informatica teorica, la linguistica formale, la teoria delle basi di dati, la verifica di sistemi, la crittografia, la semantica dei linguaggi di programmazione e le ontologie. In ciascuno di questi domini, la probabilità non è un'aggiunta esteriore, ma una condizione strutturale: la logica deve essere in grado di esprimere incertezza, di ragionare su distribuzioni, di trattare eventi rari, di comporre trasformazioni stocastiche, di verificare proprietà quantitative di sistemi dinamici.

Il primo laboratorio naturale è quello delle **basi di dati probabilistiche**. Una base di dati relazionale classica è un insieme di relazioni $R_1, \dots, R_k$ su domini $D_1, \dots, D_m$. Una base di dati probabilistica è una distribuzione di probabilità μ su un insieme di possibili istanze di tali relazioni. Il modello più studiato è quello delle tuple indipendenti: ogni tupla t ha una probabilità p_t di essere presente, e le tuple sono indipendenti. Una query Q è una formula FO (o un'espressione SQL) che definisce un insieme di possibili mondi in cui Q è vera. La probabilità della query è:

$$\Pr(Q) = \sum \{I : I \models Q\} \mu(I).$$

Nel caso di tuple indipendenti, $\mu(I) = \prod \{t \in I\} p_t \prod \{t \notin I\} (1 - p_t)$. Il problema è decidere se $\Pr(Q)$ può essere calcolata in tempo polinomiale. Risultati profondi mostrano una dicotomia: per certe classi di query FO (ad esempio congiunzioni senza join ciclici), $\Pr(Q)$ è computabile in PTIME; per altre (join ciclici, disgiunzioni complesse), il problema è #P-hard. La logica probabilistica qui si manifesta come una teoria della complessità delle formule: la struttura sintattica di Q determina la complessità della sua valutazione probabilistica. Questo è un caso paradigmatico in cui la logica probabilistica diventa una teoria della computazione probabilistica.

Un secondo laboratorio è quello dei **linguaggi di programmazione probabilistici**. Un linguaggio come Church, Anglican, WebPPL o Pyro può essere visto come un λ -calcolo esteso con operatori di campionamento e osservazione. La semantica denotazionale di tali linguaggi è data dalla monade di Giry: un termine $M : A \rightarrow B$ viene interpretato come un kernel di Markov $K_M : A \times \Sigma_B \rightarrow [0,1]$, cioè una funzione che, dato un input $a \in A$, produce una distribuzione $K_M(a, \cdot)$ su B . La composizione dei programmi è la composizione Kleisli dei kernel: se $M : A \rightarrow G(B)$ e $N : B \rightarrow G(C)$, allora $N \circ M : A \rightarrow G(C)$ è definito da:

$$(N \circ M)(a)(C') = \int_B N(b)(C') \, dM(a)(b).$$

Questa è esattamente la moltiplicazione μ della monade di Giry. La logica probabilistica entra qui come linguaggio per esprimere proprietà dei programmi: ad esempio, la probabilità che un programma termini, la distribuzione dei valori di ritorno, la probabilità che un programma soddisfi un'invariante. La logica di ordine superiore con probabilità (probabilistic type theory) è un'estensione naturale: i tipi sono spazi misurabili, i termini sono kernel, e gli operatori di integrazione e aspettazione sono primitivi del linguaggio. Il teorema di Statman sull'indcidibilità dell'uguaglianza osservazionale in λ -calcoli probabilistici mostra che la potenza espressiva di questi linguaggi è enorme: decidere se due programmi probabilistici sono equivalenti è indecidibile, perché si può codificare il comportamento di macchine di Turing probabilistiche.

Un terzo laboratorio è quello della **verifica di protocolli crittografici**. I protocolli sono sistemi dinamici con scelte probabilistiche (generazione di chiavi, nonce, scelte randomizzate). La logica probabilistica qui è la logica dei processi probabilistici: CCS probabilistico, π -calcolo probabilistico, processi Markoviani. La proprietà fondamentale è la probabilistic bisimulation: due stati s e t sono bisimili se, per ogni azione a e ogni classe di equivalenza C , la probabilità di transizione da s a C è uguale alla probabilità di transizione da t a C . Questo garantisce che s e t soddisfano le stesse formule in logiche modali probabilistiche. La verifica di anonimato, indistinguibilità, segretezza si riduce a verificare che due processi siano bisimili o quasi-bisimili. La logica probabilistica qui è uno strumento di sicurezza formale: esprime che un avversario non può distinguere due scenari con probabilità maggiore di ϵ .

Un quarto laboratorio è quello delle **ontologie probabilistiche e delle description logics probabilistiche**. Una ontologia è un insieme di assiomi del tipo $C \sqsubseteq D$, $R \sqsubseteq S$, $\exists R.C \sqsubseteq D$, ecc. In una ontologia probabilistica, ogni assioma ha una probabilità p , e si vuole calcolare la probabilità che un'asserzione $A(a)$ sia vera in un modello scelto a caso. Questo porta a logiche come Prob-ALC, Prob-SHOIQ, in cui la semantica è definita come distribuzione su modelli DL. La difficoltà è che le description logics hanno già una complessità elevata; aggiungere la probabilità porta a problemi $\#P$ -hard o addirittura EXPTIME-hard. La logica probabilistica qui diventa una teoria della ragionevolezza: quali frammenti di DL ammettono inferenza probabilistica efficiente? Quali no?

Un quinto laboratorio è quello dell'**argomentazione probabilistica**. Un framework di Dung è un grafo orientato $(A, \rightarrow)$, dove A è un insieme di argomenti e $a \rightarrow b$ significa che a attacca b . In un framework probabilistico, ogni attacco ha una probabilità $p(a \rightarrow b)$. La domanda è: qual è la probabilità che un argomento sia accettabile sotto una certa semantica (grounded, preferred, stable)? Questo porta a definire distribuzioni su sottoinsiemi di attacchi, e a calcolare la probabilità che un sottoinsieme sia un'estensione. La logica probabilistica qui è una logica di accettabilità incerta: un argomento è accettabile con probabilità $\geq r$ se, in almeno r delle possibili configurazioni, appartiene a un'estensione.

Un sesto laboratorio è quello delle **logiche infinitesimali e delle funzioni di Popper**. Qui la probabilità non è più un numero reale, ma un elemento di un campo non standard ${}^*\mathbb{R}$. Una funzione

di Popper $P(A|B)$ è definita anche quando $P(B)$ è infinitesimale. Questo permette di definire logiche condizionali in cui $\varphi > \psi$ significa che $P(\psi|\varphi)$ è “quasi 1”, cioè $1-\varepsilon$ con ε infinitesimale. La logica probabilistica qui diventa una logica della plausibilità infinitesimale: distingue eventi di probabilità zero “forte” (impossibili) da eventi di probabilità zero “debole” (infinitesimali ma possibili). Questo è cruciale per trattare paradossi come il mentitore probabilistico, dove la probabilità di verità di una frase autoreferenziale può oscillare in un intorno infinitesimale di $1/2$.

Un settimo laboratorio è quello della **teoria degli insiemi e della misurabilità universale**. Se si vuole che ogni insieme definibile in un linguaggio logico sia misurabile, occorre assumere assiomi forti come l'esistenza di un cardinale inaccessibile (Solovay). In un modello di Solovay, ogni sottoinsieme di $\mathbb{R}$ è Lebesgue misurabile. Questo permette di definire una logica probabilistica “totale” su $\mathbb{R}$: ogni formula definisce un insieme misurabile, e quindi ha una probabilità. Ma questo è possibile solo rinunciando all'assioma della scelta pieno. La logica probabilistica qui diventa una teoria dei limiti set-teorici della misurabilità.

Un ottavo laboratorio è quello dell'**apprendimento probabilistico**. La semantica PAC (Probably Approximately Correct) definisce una nozione di validità probabilistica: una formula $\varphi(x)$ è (ε, δ) -valida se

$$\Pr_{\{\text{sample} \sim D^m\}}[\Pr_{\{x \sim D\}}[\varphi(x)] \geq 1-\varepsilon] \geq 1-\delta.$$

La logica probabilistica qui diventa una logica della generalizzazione: quali formule FO hanno VC-dimension finita? Quali frammenti FO sono PAC-learnable? Risultati come il teorema di Kearns–Vazirani mostrano che FO con profondità quantificazionale limitata è PAC-learnable, mentre FO pieno no. La probabilità diventa un criterio di “robustezza apprenditiva” della logica.

Un nono laboratorio è quello delle **categorie di Markov e della monade di Giry**. Qui la probabilità è vista come struttura monadica: $G(X)$ è lo spazio delle misure su X , $\eta_X : X \rightarrow G(X)$ manda x in δ_x , $\mu_X : G(G(X)) \rightarrow G(X)$ integra misure di misure. Un kernel di Markov $f : X \rightarrow G(Y)$ è un morfismo. La logica probabilistica qui diventa una logica di invarianti categoriali: un enunciato è vero se è preservato da tutti i morfismi probabilistici. De Finetti e Aldous–Hoover diventano teoremi di fattorizzazione di morfismi simmetrici. La logica probabilistica diventa una teoria categoriale della misura.

Un decimo laboratorio è quello delle **logiche modali e temporali probabilistiche**. In PCTL, una formula come $P_{\geq r}[\varphi \cup \psi]$ significa che la probabilità che lungo un cammino ψ diventi vera prima o poi e φ rimanga vera fino a quel momento è almeno r . La semantica è definita su catene di Markov $M = (S, P, L)$, dove $P(s, t)$ è la probabilità di transizione. La probabilità di un insieme di cammini è definita come misura cilindrica:

$$\Pr_s(\{s_0, s_1, \dots \mid s_0 = s\}) = \prod_{i=0}^{\infty} P(s_i, s_{i+1}).$$

La logica probabilistica qui diventa una logica della verifica quantitativa: si vuole sapere se un sistema soddisfa una proprietà con probabilità $\geq r$. Risultati come la preservazione sotto probabilistic bisimulation mostrano che la logica è robusta rispetto a equivalenze comportamentali.

Un undicesimo laboratorio è quello delle **logiche infinitarie con probabilità**. In $L_{\omega_1 \omega, P}$, si possono esprimere proprietà come:

$$(P x \geq r) (\bigwedge_{n=1}^{\infty} \varphi_n(x)),$$

che significa che la misura dell'insieme degli x che soddisfano tutte le ϕ_n è $\geq r$. Questo permette di esprimere proprietà di convergenza, densità, limiti. Le leggi zero-uno per $L\{\omega, P\}$ mostrano che, in strutture random, molte proprietà probabilistiche si polarizzano.

Un dodicesimo laboratorio è quello delle **probabilità infinitesimali e delle logiche condizionali**. Una funzione di Popper $P(A|B)$ definita anche quando $P(B)$ è infinitesimale permette di definire logiche condizionali in cui $\phi > \psi$ significa $P(\psi|\phi) = 1 - \varepsilon$ con ε infinitesimale. Questo permette di trattare paradossi autoreferenziali e credenze revisionate.

Un tredicesimo laboratorio è quello delle **fondazioni filosofiche**. La probabilità logica non è un grado di verità fuzzy, ma una misura su uno spazio di modelli. La verità probabilistica è verità “quasi ovunque”. La logica probabilistica è la teoria della verità robusta: ciò che è vero in tutti i modelli ha probabilità 1; ciò che è vero in quasi tutti i modelli ha probabilità 1; ciò che è vero in una frazione r dei modelli ha probabilità r . La probabilità diventa una metrica sulla verità.

LABORATORIO 1 — BASI DI DATI PROBABILISTICHE

Una base di dati probabilistica è un oggetto che unisce la semantica relazionale classica con una struttura di probabilità definita su insiemi di possibili mondi. L'idea fondamentale è che un database non rappresenta più un singolo mondo deterministico, ma una distribuzione μ su un insieme di mondi possibili, ciascuno dei quali è una istanza relazionale completa. La logica probabilistica entra qui come linguaggio per esprimere proprietà dei mondi e per calcolare la probabilità che tali proprietà siano soddisfatte.

1. Struttura formale di un database probabilistico

Sia $R = \{R_1, \dots, R_k\}$ un insieme di simboli di relazione, ciascuno con una arità fissa. Un'istanza deterministica I è un insieme di tuple t tali che $t \in R_i^A$ per ogni simbolo R_i .

Definizione. Una base di dati probabilistica è una coppia (Ω, μ) , dove:

1. Ω è un insieme di istanze deterministiche I ;
2. $\mu : \Omega \rightarrow [0, 1]$ è una distribuzione di probabilità con $\sum_{I \in \Omega} \mu(I) = 1$.

Ogni $I \in \Omega$ è un possibile mondo.

Una query Q è una formula del primo ordine FO nel linguaggio relazionale R . La semantica probabilistica è:

$$\Pr(Q) = \sum_{\{I \in \Omega : I \models Q\}} \mu(I).$$

Questa è esattamente la semantica di Gaifman applicata a un dominio finito di modelli.

2. Il modello delle tuple indipendenti

Il modello più studiato è quello delle tuple indipendenti (tuple-independent databases, TID). Ogni tuple t ha una probabilità p_t di essere presente, e le tuple sono indipendenti.

Formalmente:

$\Omega = \{ I \subseteq T \}$, dove T è l'insieme di tutte le tuple possibili.

$$\mu(I) = \prod_{t \in I} p_t \prod_{t \notin I} (1 - p_t).$$

Questo è un prodotto di Bernoulli indipendenti.

Teorema 1 (Fattorizzazione della probabilità delle query congiuntive acicliche).

Sia Q una query congiuntiva senza disgiunzioni:

$$Q(x) := \exists y_1, \dots, y_m (A_1 \wedge \dots \wedge A_n),$$

dove ogni A_i è un atomo $R_j(t_1, \dots, t_r)$. Se il grafo di join di Q è aciclico, allora $\Pr(Q)$ è computabile in tempo polinomiale nel numero di tuple.

Idea della dimostrazione. Il grafo di join è un ipergrafo i cui nodi sono gli atomi A_i e gli archi rappresentano le variabili condivise. Se è aciclico, esiste un ordine di eliminazione delle variabili che non introduce cicli. La probabilità di Q può essere calcolata come:

$$\Pr(Q) = \sum_{\theta : \theta \text{ soddisfa } Q} \prod_{i=1}^n p_{A_i(\theta)}.$$

Poiché l'ipergrafo è aciclico, la somma si fattorizza in una serie di join probabilistici che possono essere calcolati tramite un algoritmo tipo Yannakakis. La complessità è polinomiale.

Teorema 2 (Dichotomia di Dalvi–Suciu).

Per ogni query congiuntiva Q , il problema di calcolare $\Pr(Q)$ in un TID è:

- in PTIME se Q è “safe” (cioè ha un piano di esecuzione senza cicli probabilistici),
- #P-hard altrimenti.

Idea della dimostrazione. La nozione di “safe plan” è definita tramite un albero di join che evita la duplicazione di variabili che generano dipendenze probabilistiche. Se Q non è safe, si può ridurre il problema del conteggio di soluzioni di una formula booleana (SAT) al calcolo di $\Pr(Q)$, mostrando #P-hardness.

Questo teorema è uno dei risultati più profondi della teoria dei database probabilistici: la complessità della valutazione probabilistica dipende unicamente dalla struttura sintattica della query.

3. Semantica logica delle query probabilistiche

Una query Q è una formula FO. La probabilità $\Pr(Q)$ è una misura logica nel senso del Volume I:

$$\mu([Q]) = \Pr(Q).$$

Il fatto che μ sia un prodotto indipendente implica che:

$$\mu([\varphi \wedge \psi]) = \mu([\varphi]) \mu([\psi]) \text{ se } \varphi \text{ e } \psi \text{ dipendono da insiemi disgiunti di tuple.}$$

Questo è un caso particolare della misura prodotto di Gaifman.

Teorema 3 (Gaifman per TID).

Sia $\varphi(x)$ una formula FO che dipende solo da un insieme finito di tuple T_φ . Allora:

$$\mu([\varphi]) = \sum \{I \subseteq T_\varphi : I \models \varphi\} \prod_{t \in I} p_t \prod_{t \notin I} (1 - p_t).$$

Dimostrazione. Segue direttamente dalla definizione di μ come prodotto indipendente e dal fatto che φ dipende solo da T_φ .

4. Query con aggregazione probabilistica

Molti linguaggi di database includono aggregazioni come COUNT, SUM, AVG. Nel contesto probabilistico, si definisce la variabile casuale:

$X_Q(I)$ = numero di tuple che soddisfano Q in I .

La distribuzione di X_Q è:

$$\Pr(X_Q = k) = \sum \{I : X_Q(I) = k\} \mu(I).$$

Teorema 4 (Distribuzione binomiale per query monadiche).

Se $Q(x)$ è una formula monadica (dipende da una sola tupla), allora X_Q è una somma di Bernoulli indipendenti, quindi:

$$X_Q \sim \text{Binomial}(n, p_Q),$$

dove $p_Q = \Pr(Q(t))$ per una tupla t generica.

Dimostrazione. Ogni tupla t contribuisce 1 a X_Q con probabilità $p_t \Pr(Q(t))$. Se tutte le tuple hanno la stessa probabilità p e $Q(t)$ è indipendente da t , allora X_Q è binomiale.

5. Query con disgiunzioni e complessità #P

Le query con disgiunzioni introducono dipendenze probabilistiche non banali.

Esempio:

$$Q := \exists x (R(x) \vee S(x)).$$

La probabilità è:

$$\Pr(Q) = 1 - \prod_{a \in D} (1 - p_{\{R(a)\}} - p_{\{S(a)\}} + p_{\{R(a)\}} p_{\{S(a)\}}).$$

Se R e S condividono variabili, la complessità esplode.

Teorema 5 (Hardness delle query disgiuntive).

Il problema di calcolare $\Pr(Q)$ per query FO con disgiunzioni è #P-hard anche se Q ha solo una disgiunzione e arità 1.

Idea della dimostrazione. Riduzione da #SAT: ogni variabile booleana è rappresentata da una tupla, e la disgiunzione simula la scelta di un assegnamento.

6. Connessione con la teoria della misura

La struttura (Ω, Σ, μ) è un caso particolare della costruzione di Scott–Krauss: ogni formula FO definisce un insieme misurabile, e μ è una misura su Σ . La differenza è che qui Σ è finita ($2^{|T|}$), quindi la misura è completamente determinata dai p_t .

Teorema 6 (Completezza di μ rispetto a FO).

Per ogni formula FO ϕ , $\mu([\phi])$ è definita e computabile come somma finita.

Dimostrazione. Ω è finito, quindi Σ è finito, quindi μ è definita su tutti gli insiemi.

7. Connessione con la complessità logica

La complessità di $\Pr(Q)$ dipende dalla struttura sintattica di Q , non dai valori p_t . Questo è un caso particolare del principio generale del Volume I: la probabilità logica è una misura su modelli, e la complessità deriva dalla struttura dell'algebra booleana generata dalle formule.

Silvano, questo è il **Laboratorio 1** sviluppato come un capitolo completo, con teoremi e dimostrazioni discorsive.

Se vuoi, procedo ora con il **Laboratorio 2: Linguaggi di programmazione probabilistici**, con la stessa profondità.

Modifica in una pagina

Hai detto

Si, ma con meno schematicità

Copilot ha detto

Perfetto, Silvano. Allora **riprendo il Laboratorio 1** e lo riscrivo **in forma completamente discorsiva**, senza schemi, senza elenchi, senza titoli rigidi, come un capitolo di un libro maturo: fluido, continuo, ma con la stessa profondità tecnica, con teoremi e dimostrazioni integrati nella narrazione, come farebbe un matematico che racconta la teoria mentre la sviluppa.

Procedo ora.

Quando si parla di basi di dati probabilistiche, la prima cosa da chiarire è che non si tratta di un'estensione cosmetica del modello relazionale, ma di un vero e proprio cambio di paradigma: un database non è più un singolo mondo, ma una distribuzione di probabilità su un insieme di mondi possibili. Ogni mondo è un'istanza completa delle relazioni, e la probabilità misura quanto quel mondo è plausibile. In questo senso, una base di dati probabilistica è un caso particolare della semantica di Gaifman: un insieme di modelli, ciascuno con una probabilità, e una logica che valuta la probabilità che una formula sia vera in un modello scelto a caso.

Per rendere tutto questo rigoroso, si parte da un insieme di simboli di relazione $R_1, \dots, R_k$, ciascuno con un'arietà fissata. Un mondo possibile è un'istanza deterministica I , cioè un insieme di tuple che popolano le relazioni. Una base di dati probabilistica è una distribuzione μ su un insieme Ω di tali istanze. Una query Q è una formula del primo ordine nel linguaggio relazionale, e la sua probabilità è semplicemente la misura dell'insieme dei mondi in cui Q è vera: $\Pr(Q) = \sum_{\{I \in \Omega, I \models Q\}} \mu(I)$. Questa è esattamente la definizione di $\mu([Q])$ nel senso del Volume I.

Il modello più studiato è quello delle tuple indipendenti. Ogni tupla t ha una probabilità p_t di essere presente, e le tuple sono indipendenti. In questo caso, Ω è l'insieme di tutti i sottoinsiemi di T , l'insieme delle tuple possibili, e $\mu(I) = \prod_{\{t \in I\}} p_t \prod_{\{t \notin I\}} (1 - p_t)$. Questa è una misura prodotto, e quindi è un caso particolare della costruzione di Scott–Krauss: ogni tupla è una variabile casuale indipendente, e la probabilità di un mondo è la probabilità che tutte le variabili assumano i valori indicati da quel mondo.

La domanda naturale è: come si calcola $\Pr(Q)$? Per una query congiuntiva semplice, come $\exists x, y (R(x, y) \wedge S(y))$, la probabilità è la somma delle probabilità dei mondi in cui esiste una coppia (a, b) tale che $R(a, b)$ e $S(b)$ sono vere. Ma questa somma è, in generale, esponenziale nel numero di tuple. La struttura logica della query determina se la probabilità è calcolabile in tempo polinomiale o se è intrinsecamente difficile.

Qui entra in gioco un risultato fondamentale, che nella letteratura è noto come la dicotomia di Dalvi–Suciu. L'idea è che una query congiuntiva Q può essere rappresentata come un ipergrafo di join: ogni atomo è un nodo, e due nodi sono collegati se condividono variabili. Se questo ipergrafo è aciclico, allora la probabilità di Q può essere calcolata in tempo polinomiale. Se è ciclico, il problema diventa $\#P$ -hard. Questo è un risultato sorprendente: la complessità non dipende dai valori delle probabilità p_t , ma solo dalla struttura sintattica della query.

Per capire perché l'assenza di cicli rende tutto facile, si può osservare che, se il grafo di join è aciclico, esiste un ordine di eliminazione delle variabili che non introduce dipendenze probabilistiche indesiderate. In questo caso, la probabilità di Q può essere calcolata con un algoritmo simile a quello di Yannakakis per il join aciclico: si eliminano le variabili una alla volta, mantenendo una rappresentazione compatta delle distribuzioni intermedie. Il risultato è un algoritmo polinomiale.

Se invece il grafo di join ha un ciclo, la situazione cambia radicalmente. La presenza di un ciclo introduce dipendenze probabilistiche che non possono essere eliminate localmente. In questo caso, si può dimostrare che calcolare $\Pr(Q)$ è $\#P$ -hard, anche se Q è molto semplice. La dimostrazione tipica consiste nel ridurre il problema del conteggio delle soluzioni di una formula booleana (il problema $\#SAT$) al calcolo di $\Pr(Q)$. Si costruisce un database probabilistico in cui ogni variabile booleana è rappresentata da una tupla, e la query Q simula la formula booleana. La probabilità che Q sia vera è proporzionale al numero di assegnamenti soddisfacenti della formula. Poiché $\#SAT$ è $\#P$ -hard, anche $\Pr(Q)$ lo è.

Questo risultato è un esempio perfetto del principio generale del Volume I: la probabilità logica è una misura su un'algebra booleana generata dalle formule, e la complessità deriva dalla struttura di questa algebra. Nel caso delle query congiuntive, l'algebra è generata dagli atomi $R_i(t)$, e la presenza di cicli corrisponde a una struttura booleana non decomponibile.

Un altro aspetto importante è la semantica delle query con aggregazione. Se $Q(x)$ è una formula monadica, cioè dipende da una sola tupla, allora la variabile casuale $X_Q(I)$ = numero di tuple che soddisfano Q in I è una somma di Bernoulli indipendenti. In questo caso, X_Q ha distribuzione binomiale: $X_Q \sim \text{Binomial}(n, p_Q)$, dove p_Q è la probabilità che una tupla generica soddisfi Q .

Questo è un caso particolare del teorema di de Finetti: una somma di variabili indipendenti identicamente distribuite ha una distribuzione determinata dalla misura prodotto.

Le query con disgiunzioni introducono invece dipendenze non banali. Una query come $\exists x (R(x) \vee S(x))$ ha una probabilità che non si fattorizza facilmente. Se R e S sono indipendenti, la probabilità che una tupla a soddisfi $R(a) \vee S(a)$ è $p_R(a) + p_S(a) - p_R(a)p_S(a)$. Ma se R e S condividono variabili, la situazione si complica. Si può dimostrare che calcolare $\Pr(Q)$ per query con disgiunzioni è #P-hard anche se Q ha una sola disgiunzione e arità 1. La dimostrazione è una riduzione da #SAT: si rappresentano le variabili booleane come tuple, e la disgiunzione simula la scelta di un assegnamento.

Tutto questo mostra che la teoria delle basi di dati probabilistiche è, in realtà, una teoria della misura su insiemi di modelli FO. La misura μ è un caso particolare della misura di Scott–Krauss: ogni formula FO definisce un insieme misurabile, e μ è una misura su Σ_L . La differenza è che qui Σ_L è finita, quindi la misura è completamente determinata dai p_t . Ma la struttura logica è la stessa: la probabilità di una formula è la misura dell'insieme dei modelli che la soddisfano.

Questa prospettiva permette di vedere i database probabilistici come un caso concreto della teoria generale del Volume I. La misura μ è una misura prodotto, quindi è un caso particolare della misura di Gaifman. Le query FO sono formule, quindi definiscono insiemi misurabili. La complessità della valutazione probabilistica deriva dalla struttura dell'algebra booleana generata dagli atomi. E la dicotomia di Dalvi–Suciu è un caso particolare del principio generale: la complessità logica è determinata dalla struttura sintattica delle formule.

LABORATORIO 2 — Linguaggi di programmazione probabilistici

Quando si parla di linguaggi di programmazione probabilistici, si entra in un territorio in cui la logica, la teoria della misura e la semantica dei linguaggi si intrecciano in modo estremamente profondo. Un programma probabilistico non è semplicemente un algoritmo che usa numeri casuali: è un oggetto semantico che denota una **trasformazione stocastica**, cioè una funzione che, dato un input, produce una distribuzione di probabilità sugli output. La logica probabilistica entra qui come linguaggio per esprimere proprietà quantitative dei programmi: la probabilità che un programma termini, la distribuzione dei valori di ritorno, la probabilità che un invariante venga violato, la probabilità che un programma soddisfi una specifica temporale.

Per capire tutto questo, bisogna partire dalla semantica denotazionale. In un linguaggio deterministico, un termine $M : A \rightarrow B$ denota una funzione $f_M : A \rightarrow B$. In un linguaggio probabilistico, un termine $M : A \rightarrow B$ denota un **kernel di Markov**, cioè una funzione

$$K_M : A \times \Sigma_B \rightarrow [0,1]$$

tale che, per ogni $a \in A$, la funzione $K_M(a, \cdot)$ è una misura di probabilità su B , e per ogni insieme misurabile $E \subseteq B$, la funzione $a \mapsto K_M(a, E)$ è misurabile. In altre parole, $K_M(a, \cdot)$ è la distribuzione degli output quando il programma M viene eseguito con input a .

Questa struttura è esattamente quella della **monade di Giry**, che nel Volume I abbiamo introdotto come il funtore G che manda uno spazio misurabile X nello spazio $G(X)$ delle misure di probabilità su X . La semantica di un programma probabilistico è un morfismo nella categoria di Kleisli della monade di Giry:

$$[[M]] : A \rightarrow G(B).$$

La composizione dei programmi è la composizione Kleisli:

$$\llbracket N \circ M \rrbracket = \mu_B \circ G(\llbracket N \rrbracket) \circ \llbracket M \rrbracket,$$

dove $\mu_B : G(G(B)) \rightarrow G(B)$ è la moltiplicazione della monade, definita da

$$\mu_B(\Pi)(E) = \int \{v \in G(B)\} v(E) d\Pi(v).$$

Questa formula è il cuore della semantica probabilistica: comporre due programmi probabilistici significa integrare le distribuzioni generate dal primo programma rispetto alle distribuzioni generate dal secondo.

Per rendere tutto questo più concreto, consideriamo un linguaggio minimale con due costrutti probabilistici:

1. **sample(d)**, che campiona un valore da una distribuzione d;
2. **observe(ϕ)**, che condiziona l'esecuzione al fatto che ϕ sia vera.

Il costrutto **sample(d)** denota un kernel $K_{\text{sample}} : 1 \rightarrow G(X)$, dove 1 è il tipo unitario, e $K_{\text{sample}}(*, E) = d(E)$. Il costrutto **observe(ϕ)** è più delicato: se $\phi(x)$ è una proprietà misurabile, allora **observe(ϕ)** denota un kernel che, dato un input x, restituisce la misura condizionata $\delta_x | \phi$, cioè la delta di Dirac su x se $\phi(x)$ è vera, e la misura zero altrimenti. In un linguaggio realistico, **observe** è usato per modellare l'evidenza in un modello bayesiano: si campiona un valore latente, si osserva un dato, e si condiziona la distribuzione latente al dato osservato.

La semantica di un programma probabilistico con **observe** è quindi una **misura condizionata**, e questo introduce immediatamente problemi di definibilità: la misura condizionata esiste solo se l'evento osservato ha probabilità positiva. Per questo motivo, molti linguaggi probabilistici adottano una semantica non normalizzata: **observe(ϕ)** moltiplica la densità per l'indicatore di ϕ , e la normalizzazione viene effettuata solo alla fine.

Per formalizzare tutto questo, si introduce la nozione di **kernel subprobabilistico**: una funzione

$$K : A \times \Sigma_B \rightarrow [0,1]$$

tale che $K(a, B) \leq 1$. La semantica di **observe(ϕ)** è un kernel subprobabilistico:

$$K_{\text{obs}}(a, E) = 1_E(a) \cdot 1_{\{\phi(a)\}}.$$

La normalizzazione finale è ottenuta dividendo per la massa totale.

Questa semantica è perfettamente coerente con la teoria della misura: un kernel subprobabilistico è una misura di probabilità “difettosa”, e la normalizzazione è una divisione per la massa totale. La logica probabilistica entra qui come linguaggio per esprimere proprietà di questi kernel: ad esempio, la probabilità che un programma termini è la massa totale della misura generata dal programma.

Un risultato fondamentale in questo contesto è il **teorema di adeguatezza probabilistica**: la semantica denotazionale coincide con la semantica operativa. La semantica operativa definisce la distribuzione degli output come limite delle distribuzioni generate dalle esecuzioni finite del programma. La semantica denotazionale definisce la distribuzione come kernel di Markov. Il teorema afferma che queste due definizioni coincidono.

Formalmente, se $\llbracket M \rrbracket_{\text{op}}$ è la distribuzione operazionale e $\llbracket M \rrbracket_{\text{den}}$ è la distribuzione denotazionale, allora:

$$\llbracket M \rrbracket_{\text{op}} = \llbracket M \rrbracket_{\text{den}}.$$

La dimostrazione è un'analisi di convergenza: si mostra che la semantica operazionale genera una misura crescente (in senso di misura subprobabilistica), e che il limite di questa misura è esattamente la misura definita dal kernel denotazionale. Questo richiede l'uso del teorema della convergenza monotona e della σ -additività della misura.

Un altro risultato fondamentale è l'indecidibilità dell'equivalenza osservazionale. Due programmi M e N sono osservazionalmente equivalenti se, per ogni contesto $C[\cdot]$, la distribuzione degli output di $C[M]$ è uguale a quella di $C[N]$. Si può dimostrare che decidere se due programmi probabilistici sono equivalenti è indecidibile. La dimostrazione consiste nel codificare una macchina di Turing probabilistica in un programma probabilistico, e nel mostrare che l'equivalenza di due programmi corrisponde all'uguaglianza delle distribuzioni generate da due macchine di Turing. Poiché l'uguaglianza delle distribuzioni generate da macchine di Turing è indecidibile, anche l'equivalenza osservazionale lo è.

Questo risultato mostra che i linguaggi probabilistici sono estremamente potenti: possono simulare qualsiasi processo stocastico computabile, e la loro semantica è ricca quanto la teoria della misura stessa.

Un aspetto particolarmente interessante è la connessione con la logica probabilistica. Una proprietà di un programma probabilistico può essere espressa come una formula in una logica modale probabilistica: ad esempio, la probabilità che un programma termini può essere espressa come $P_{\geq r}(\text{terminate})$. La semantica di questa formula è definita rispetto al kernel denotazionale del programma. La logica probabilistica diventa quindi un linguaggio per verificare proprietà quantitative dei programmi.

Un esempio classico è la logica PCTL (Probabilistic Computation Tree Logic), che permette di esprimere proprietà come “la probabilità che il programma raggiunga uno stato di errore entro 10 passi è ≤ 0.01 ”. La semantica di PCTL è definita su catene di Markov, e un programma probabilistico può essere visto come una catena di Markov generata dinamicamente.

La connessione con la teoria delle categorie è ancora più profonda. La monade di Giry fornisce una struttura monadica che permette di comporre kernel di Markov. Le categorie di Markov forniscono un linguaggio diagrammatico per ragionare su trasformazioni probabilistiche. In questo linguaggio, un programma probabilistico è un morfismo, e la composizione dei programmi è la composizione dei morfismi. La logica probabilistica diventa una teoria degli invarianti categoriali: un enunciato è vero se è preservato da tutti i morfismi probabilistici.

In questo senso, i linguaggi di programmazione probabilistici sono un laboratorio perfetto per la logica probabilistica: ogni programma è una misura, ogni costrutto è un kernel, ogni composizione è un'integrazione, ogni proprietà è una formula probabilistica.

LABORATORIO 3 — Verifica di protocolli crittografici probabilistici

Quando si entra nel dominio della crittografia formale, la probabilità non è un optional: è la sostanza stessa del comportamento dei protocolli. Ogni protocollo moderno — dallo scambio di chiavi Diffie–Hellman ai sistemi di autenticazione challenge–response, fino ai protocolli di anonimato come Tor — è costruito su scelte casuali, nonce, generazione di chiavi, rumore crittografico, distribuzioni uniformi su gruppi, e avversari che operano con probabilità di successo misurabili.

La logica probabilistica diventa quindi il linguaggio naturale per esprimere proprietà come:

- “la probabilità che un avversario distingua due scenari è $\leq \epsilon$ ”;
- “la probabilità che un protocollo riveli la chiave è trascurabile”;
- “la probabilità che un attaccante impersoni un utente è $\leq 2^{-128}$ ”;
- “due processi sono indistinguibili per ogni avversario polinomiale”.

Per formalizzare tutto questo, bisogna partire da una semantica operativa probabilistica dei protocolli. Un protocollo è un processo che evolve nel tempo, con scelte probabilistiche interne e interazioni con un avversario. La semantica naturale è quella dei **processi probabilistici**, cioè sistemi di transizione etichettati con probabilità.

Formalmente, un processo probabilistico è una tripla:

$$P = (S, \rightarrow, L),$$

dove:

- S è un insieme di stati;
- $\rightarrow \subseteq S \times \text{Dist}(S)$ è una relazione di transizione probabilistica, dove $\text{Dist}(S)$ è l'insieme delle distribuzioni di probabilità su S ;
- $L : S \rightarrow \text{Obs}$ è una funzione di osservazione, che assegna a ogni stato ciò che un avversario può osservare.

Una transizione $s \rightarrow \mu$ significa che, dallo stato s , il processo evolve in uno stato t con probabilità $\mu(t)$. Questo è un caso particolare dei kernel di Markov del Volume II: ogni stato s induce una distribuzione $K(s, \cdot)$ sugli stati successivi.

La proprietà fondamentale che si vuole verificare è l'**indistinguibilità probabilistica**. Due processi P e Q sono indistinguibili se nessun avversario può distinguerli con probabilità significativamente maggiore di $1/2$. Per formalizzare questo, si introduce la nozione di **probabilistic bisimulation**, che è la versione probabilistica della bisimulazione classica dei processi.

Definizione. Una relazione $R \subseteq S_P \times S_Q$ è una probabilistic bisimulation se, per ogni $(s, t) \in R$:

1. $L_P(s) = L_Q(t)$;
2. per ogni transizione $s \rightarrow \mu$ in P , esiste una transizione $t \rightarrow \nu$ in Q tale che, per ogni classe C di R , $\mu(C) = \nu(C)$;
3. simmetricamente, per ogni transizione $t \rightarrow \nu$ in Q , esiste una transizione $s \rightarrow \mu$ in P con la stessa proprietà.

Due stati s e t sono probabilisticamente bisimili se esiste una relazione R che li collega.

Questa definizione è un raffinamento della nozione di equivalenza comportamentale: due stati sono bisimili se hanno la stessa struttura probabilistica di transizione, modulo rinominazione degli stati.

Il teorema fondamentale è che la probabilistic bisimulation è esattamente la relazione di equivalenza per le logiche modali probabilistiche.

Teorema. Siano P e Q due processi probabilistici. Allora:

$s \sim t$ (bisimili) se e solo se per ogni formula φ della logica modale probabilistica, $P, s \models \varphi \Leftrightarrow Q, t \models \varphi$.

La logica modale probabilistica ha formule del tipo:

- $[a]_{\geq p} \varphi$, che significa: “dopo un’azione a , la probabilità di raggiungere uno stato che soddisfa φ è $\geq p$ ”;
- combinazioni booleane di tali formule.

La dimostrazione è un adattamento del classico teorema di Hennessy–Milner, con la differenza che le transizioni sono probabilistiche e quindi le formule devono parlare di misure, non solo di possibilità.

Questo teorema è cruciale per la verifica crittografica: se due processi sono bisimili, allora nessun avversario può distinguerli osservando le loro uscite. In altre parole, la probabilistic bisimulation è una nozione formale di **indistinguibilità crittografica**.

Ma la crittografia moderna richiede una nozione più forte: l’indistinguibilità rispetto a un avversario polinomiale. Questo porta alla definizione di **indistinguibilità computazionale**.

Definizione. Due distribuzioni D_0 e D_1 sono computazionalmente indistinguibili se, per ogni algoritmo probabilistico A che gira in tempo polinomiale, vale:

$$|\Pr[A(D_0)=1] - \Pr[A(D_1)=1]| \leq \varepsilon(n),$$

dove $\varepsilon(n)$ è una funzione trascurabile.

Questa definizione è la base della crittografia moderna: un cifrario è sicuro se il ciphertext è computazionalmente indistinguibile da un ciphertext casuale.

La logica probabilistica entra qui come linguaggio per esprimere indistinguibilità: una formula φ può esprimere che due processi sono indistinguibili per ogni avversario polinomiale. La semantica è definita rispetto a un modello in cui gli avversari sono processi probabilistici con risorse limitate.

Un risultato fondamentale è il **teorema di composizione**: se due protocolli sono indistinguibili, allora la loro composizione è indistinguibile. Formalmente:

Teorema (Composizione sequenziale). Se $P_0 \approx P_1$ e $Q_0 \approx Q_1$, allora $Q_0 \circ P_0 \approx Q_1 \circ P_1$.

La dimostrazione usa la linearità della probabilità e la monotonicità della distanza statistica:

$$\Delta(\mu, \nu) = \sup_{\{E\}} |\mu(E) - \nu(E)|.$$

Se $\Delta(P_0, P_1) \leq \varepsilon$ e $\Delta(Q_0, Q_1) \leq \varepsilon$, allora $\Delta(Q_0 \circ P_0, Q_1 \circ P_1) \leq 2\varepsilon$.

Questo è un caso particolare della proprietà di non espansività dei kernel di Markov: la distanza statistica non aumenta sotto trasformazioni stocastiche.

Un altro risultato fondamentale è il **teorema di equivalenza tra bisimulazione e distanza statistica**. Se due processi sono bisimili, allora la distanza statistica tra le distribuzioni generate è zero. Se non sono bisimili, la distanza statistica è positiva.

Formalmente:

Teorema. Siano P e Q due processi probabilistici. Allora:

$$s \sim t \Leftrightarrow \Delta(\mu_s, \mu_t) = 0,$$

dove μ_s è la distribuzione dei cammini generati da P a partire da s .

La dimostrazione è un'analisi di coinduzione: se s e t sono bisimili, allora ogni passo produce distribuzioni identiche modulo la relazione, e quindi la distribuzione dei cammini è identica.

Questo teorema collega direttamente la logica modale probabilistica con la teoria della misura: la bisimulazione è una condizione di uguaglianza di misure.

La verifica dei protocolli crittografici si basa su questi strumenti: si modella il protocollo come un processo probabilistico, si modella l'avversario come un processo che interagisce con il protocollo, e si verifica che il protocollo e un protocollo ideale sono bisimili o indistinguibili.

Un esempio classico è il protocollo di anonimato: un utente U invia un messaggio m attraverso un protocollo P . Si vuole dimostrare che un avversario non può distinguere se il messaggio è stato inviato da U o da un altro utente V . Si modella $P(U, m)$ e $P(V, m)$ come due processi probabilistici, e si dimostra che sono bisimili modulo rinominazione degli utenti. Questo implica che nessun avversario può distinguerli.

Un altro esempio è la sicurezza IND-CPA dei cifrari: si modella il cifrario come un processo probabilistico che, dato un messaggio m , produce un ciphertext c . Si dimostra che il processo che cifra m_0 e quello che cifra m_1 sono computazionalmente indistinguibili. Questo è un caso particolare della bisimulazione computazionale.

In tutti questi casi, la logica probabilistica è il linguaggio che permette di esprimere e verificare proprietà quantitative dei protocolli: probabilità di successo dell'avversario, distanza statistica tra processi, indistinguibilità computazionale, sicurezza compositiva.

LABORATORIO 4 — Ontologie probabilistiche e Description Logics probabilistiche

Quando si passa dalle logiche del primo ordine alle logiche descrittive, si entra in un territorio in cui la struttura concettuale è più rigida, più modulare, più orientata alla rappresentazione della conoscenza che alla modellazione arbitraria di strutture. Le description logics (DL) sono nate per fornire un linguaggio formale per ontologie, tassonomie, gerarchie concettuali, relazioni tra classi e individui. Sono la base teorica di OWL, e quindi dell'intero ecosistema del Web Semantico.

L'aggiunta della probabilità a questo quadro non è un semplice arricchimento: è un cambiamento strutturale. Una ontologia probabilistica non è più un insieme di assiomi che devono essere veri in tutti i modelli, ma un insieme di assiomi ciascuno dei quali ha una **probabilità di essere vero**. Questo porta a una semantica radicalmente diversa: non si cerca più un modello che soddisfi tutti gli assiomi,

ma una **distribuzione di probabilità sui modelli** che soddisfi gli assiomi con le probabilità specificate.

Per capire questo, bisogna partire dalla semantica classica delle DL. In ALC, il frammento base delle description logics, i concetti sono costruiti da:

- concetti atomici A ,
- negazione $\neg C$,
- congiunzione $C \sqcap D$,
- esistenziale $\exists R.C$,
- universale $\forall R.C$.

Un modello $M = (\Delta^M, \cdot^M)$ interpreta ogni concetto come un sottoinsieme di Δ^M e ogni ruolo R come una relazione binaria su Δ^M . Un assioma $C \sqsubseteq D$ è vero in M se $C^M \subseteq D^M$.

In una ontologia probabilistica, ogni assioma ha una probabilità p . Ad esempio:

$$\begin{array}{ll} (1) \text{ Bird } \sqsubseteq \text{ Fly} & p = 0.95 \\ p = 1 \quad (3) \text{ Penguin } \sqsubseteq \neg \text{Fly} & p = 1 \end{array} \quad (2) \text{ Penguin } \sqsubseteq \text{ Bird}$$

Questa ontologia esprime che “quasi tutti gli uccelli volano”, ma “tutti i pinguini non volano”. La semantica classica non può rappresentare questa situazione: gli assiomi (1) e (3) sono incompatibili se $\text{Penguin} \sqsubseteq \text{Bird}$ è vero. La semantica probabilistica invece permette di dire che (1) è vero nel 95% dei modelli, mentre (3) è vero nel 100%.

Formalmente, una ontologia probabilistica è una distribuzione μ su un insieme di modelli M tali che:

$$\mu(\{M \mid M \models C \sqsubseteq D\}) = p.$$

Questa è esattamente la semantica di Gaifman applicata alle description logics: ogni assioma è una formula FO, e la probabilità è la misura dell'insieme dei modelli che la soddisfano.

Il problema fondamentale è calcolare la probabilità che un'asserzione $A(a)$ sia vera. Questa è la **probabilità di inferenza**:

$$\Pr(A(a)) = \mu(\{M \mid M \models A(a)\}).$$

Il calcolo di questa probabilità è il cuore della logica descrittiva probabilistica.

Per rendere tutto questo computabile, si introduce un modello di base: il **modello delle scelte indipendenti**. Ogni assioma $C \sqsubseteq D$ con probabilità p è trattato come una variabile booleana indipendente che può essere vera con probabilità p e falsa con probabilità $1-p$. Ogni possibile mondo è una scelta di quali assiomi sono veri. Per ogni mondo ω , si costruisce una ontologia deterministica O_ω contenente gli assiomi scelti. Se O_ω è consistente, si valuta se $O_\omega \models A(a)$. La probabilità è:

$$\Pr(A(a)) = \sum\{\omega : O_\omega \models A(a)\} \mu(\omega).$$

Questa è una somma esponenziale, e quindi il problema è in generale #P-hard.

Il risultato più importante in questo contesto è la **dichotomia di probabilistic DL inference**: per alcuni frammenti di DL, il problema è in PTIME; per altri, è #P-hard.

Per esempio, nel frammento EL (che permette solo congiunzione e esistenziale), la probabilità di inferenza può essere calcolata in tempo polinomiale. Questo è sorprendente: EL è molto espressiva, ma la sua struttura monotona permette algoritmi efficienti.

Il motivo è che, in EL, l'inferenza è monotona: se un assioma è vero in un mondo, aggiungerlo non può rendere falsa un'inferenza. Questo permette di usare algoritmi di tipo forward-chaining e di calcolare la probabilità come prodotto di probabilità indipendenti.

Invece, in ALC, la presenza della negazione rompe la monotonicità. Un assioma può rendere inconsistente un mondo, e quindi la probabilità deve essere calcolata escludendo i mondi inconsistenti. Questo introduce dipendenze complesse e porta alla #P-hardness.

Un risultato formale è il seguente.

Teorema. Il problema di calcolare $\Pr(A(a))$ in una ontologia probabilistica ALC è #P-hard anche se l'ontologia contiene solo assiomi del tipo $C \sqsubseteq D$ con probabilità p .

Idea della dimostrazione. Si riduce il problema del conteggio delle soluzioni di una formula booleana (il problema #SAT) al calcolo della probabilità di un concetto. Ogni variabile booleana è rappresentata da un assioma probabilistico, e la formula booleana è rappresentata da un concetto complesso. La probabilità che il concetto sia soddisfatto è proporzionale al numero di assegnamenti soddisfacenti.

Questo risultato mostra che la logica descrittiva probabilistica è, in generale, computazionalmente difficile. Ma esistono frammenti più semplici.

Un altro approccio è quello delle **ontologie probabilistiche con distribuzioni esplicite**. Invece di assegnare probabilità agli assiomi, si assegnano distribuzioni ai concetti. Ad esempio:

$$\Pr(\text{Fly} \mid \text{Bird}) = 0.95.$$

Questa è una forma di logica probabilistica condizionale. La semantica è definita tramite misure su Δ^M : per ogni individuo x , la probabilità che x appartenga a Fly è 0.95 volte la probabilità che appartenga a Bird. Questo porta a una semantica di tipo Markov logic: ogni assioma è un vincolo probabilistico su insiemi di individui.

Un risultato fondamentale è che questa semantica è equivalente a una Markov Logic Network (MLN) quando i concetti sono trattati come predicati unari. In questo caso, la probabilità di un modello M è:

$$\Pr(M) \propto \exp\left(\sum_i w_i N_i(M)\right),$$

dove $N_i(M)$ è il numero di istanze dell'assioma i in M , e w_i è un peso logaritmico legato alla probabilità.

Questo collega direttamente le description logics probabilistiche con la logica statistica di Markov.

Un altro approccio ancora è quello delle **ontologie bayesiane**, in cui i concetti sono variabili casuali e le relazioni tra concetti sono archi di un grafo bayesiano. In questo caso, la probabilità di un concetto è calcolata tramite inferenza bayesiana. Questo è un caso particolare della semantica di Carnap: ogni concetto ha una distribuzione a priori, e le osservazioni aggiornano la distribuzione.

In tutti questi casi, la logica probabilistica è il linguaggio che permette di esprimere e calcolare la probabilità che un concetto sia vero, che un individuo appartenga a una classe, che un ruolo sia soddisfatto, che un assioma sia valido.

La difficoltà computazionale deriva dalla struttura logica: la presenza della negazione, dell'universale, della disgiunzione rende il problema difficile. La monotonicità di EL rende il problema facile. La probabilità è quindi un criterio di complessità logica: frammenti monotoni sono computabili, frammenti non monotoni sono difficili.

LABORATORIO 5 — Argomentazione probabilistica

L'argomentazione formale nasce come tentativo di modellare il ragionamento non monotono: situazioni in cui nuove informazioni possono invalidare conclusioni precedenti. Il quadro classico è quello dei **framework di Dung**, in cui un insieme di argomenti A è accompagnato da una relazione di attacco $\rightarrow \subseteq A \times A$. Un argomento a attacca b se, in qualche senso, a fornisce una ragione per rifiutare b . Le estensioni — grounded, preferred, stable — sono insiemi di argomenti che si sostengono a vicenda e respingono gli attacchi.

Tutto questo è deterministico: o un attacco esiste, o non esiste; o un argomento è accettabile, o non lo è. Ma nella pratica, gli attacchi sono spesso incerti: un argomento può attaccarne un altro con una certa probabilità, oppure la presenza stessa di un argomento può essere incerta. Questo porta naturalmente alla **argomentazione probabilistica**, in cui la struttura di attacco è soggetta a incertezza, e l'accettabilità di un argomento diventa una variabile casuale.

Per formalizzare tutto questo, si parte da un framework di Dung probabilistico, che è una coppia (A, P) , dove A è un insieme finito di argomenti e P è una distribuzione di probabilità su sottoinsiemi di attacchi. Ogni possibile mondo ω è un sottoinsieme $E_\omega \subseteq A \times A$, che rappresenta gli attacchi presenti in quel mondo. La probabilità $\mu(\omega)$ è la probabilità che quella configurazione di attacchi sia la vera struttura argomentativa.

In ogni mondo ω , si può calcolare l'estensione grounded G_ω , l'estensione preferred P_ω , l'estensione stable S_ω , ecc. L'accettabilità di un argomento a diventa una variabile casuale:

$$X_a(\omega) = 1 \text{ se } a \in G_\omega, X_a(\omega) = 0 \text{ altrimenti.}$$

La probabilità che a sia accettabile è:

$$\Pr(a \text{ accettabile}) = \sum \{ \omega : a \in G_\omega \} \mu(\omega).$$

Questa è esattamente la semantica di Gaifman applicata ai framework di Dung: ogni estensione è un insieme di modelli, e la probabilità è la misura dell'insieme dei mondi in cui l'argomento è accettabile.

Il problema fondamentale è calcolare questa probabilità. Se gli attacchi sono indipendenti, ogni attacco $(a \rightarrow b)$ è presente con probabilità $p_{\{ab\}}$, e assente con probabilità $1 - p_{\{ab\}}$. In questo caso, μ è una misura prodotto:

$$\mu(\omega) = \prod \{ (a \rightarrow b) \in \omega \} p_{\{ab\}} \prod \{ (a \rightarrow b) \notin \omega \} (1 - p_{\{ab\}}).$$

Il numero di mondi è $2^{|A \times A|}$, quindi la somma è esponenziale. La complessità deriva dalla struttura logica dell'accettabilità: l'estensione grounded è definita come il punto fisso minimo dell'operatore di accettabilità, e questo punto fisso dipende in modo non monotono dagli attacchi.

Per capire la complessità, consideriamo un esempio semplice. Sia $A = \{a, b, c\}$, con attacchi probabilistici:

$a \rightarrow b$ con probabilità p , $b \rightarrow c$ con probabilità q .

In un mondo ω , l'estensione grounded è:

- $\{a, c\}$ se $a \rightarrow b$ è presente e $b \rightarrow c$ è assente;
- $\{a\}$ se $a \rightarrow b$ è presente e $b \rightarrow c$ è presente;
- $\{a, b, c\}$ se $a \rightarrow b$ è assente.

La probabilità che c sia accettabile è:

$$\Pr(c \text{ accettabile}) = (1-p) + p(1-q).$$

Questa è una formula semplice, ma già mostra la natura non lineare dell'accettabilità: la presenza di un attacco può escludere un argomento, ma la presenza di un altro attacco può riabilitarlo.

Il risultato generale è che il calcolo della probabilità di accettabilità è #P-hard. La dimostrazione consiste nel ridurre il problema del conteggio delle soluzioni di una formula booleana al calcolo della probabilità di accettabilità. Ogni variabile booleana è rappresentata da un attacco probabilistico, e la formula booleana è rappresentata da un argomento complesso. La probabilità che l'argomento sia accettabile è proporzionale al numero di assegnamenti soddisfacenti.

Formalmente:

Teorema. Il problema di calcolare $\Pr(a \text{ accettabile})$ in un framework di Dung probabilistico con attacchi indipendenti è #P-hard, anche se ogni attacco ha probabilità $1/2$.

Idea della dimostrazione. Si costruisce un framework in cui ogni variabile booleana x_i è rappresentata da un attacco probabilistico ($u_i \rightarrow v_i$). La formula booleana ϕ è rappresentata da un argomento t che è attaccato da argomenti che rappresentano le clausole. La probabilità che t sia accettabile è proporzionale al numero di assegnamenti che soddisfano ϕ . Poiché il conteggio delle soluzioni di ϕ è #P-hard, anche $\Pr(t \text{ accettabile})$ lo è.

Questo risultato mostra che l'argomentazione probabilistica è computazionalmente difficile in generale. Ma esistono frammenti più semplici.

Un caso importante è quello dei framework **stratificati**, in cui gli attacchi formano un DAG. In questo caso, l'estensione grounded può essere calcolata in tempo polinomiale, e la probabilità di accettabilità può essere calcolata con un algoritmo dinamico. L'idea è che, in un DAG, l'accettabilità di un argomento dipende solo dagli argomenti che lo precedono nel DAG, e quindi si può calcolare la probabilità in modo bottom-up.

Un altro caso importante è quello dei framework **monotoni**, in cui l'aggiunta di un attacco non può rendere accettabile un argomento che prima non lo era. In questo caso, la probabilità di accettabilità è monotona rispetto ai $p_{\{ab\}}$, e può essere calcolata come prodotto di probabilità indipendenti.

Un risultato interessante è che la probabilità di accettabilità può essere vista come una misura di robustezza argomentativa: un argomento è robusto se rimane accettabile in molti mondi. Questo collega l'argomentazione probabilistica alla teoria della misura: la probabilità è una misura della stabilità dell'argomento rispetto alle perturbazioni della struttura.

Un altro collegamento profondo è con la teoria dei giochi. L'estensione grounded può essere vista come l'equilibrio di un gioco in cui gli argomenti competono per essere accettati. La probabilità di accettabilità è la probabilità che un argomento vinca il gioco in un mondo casuale. Questo collega l'argomentazione probabilistica alla teoria dei giochi bayesiana.

Infine, c'è un collegamento con la logica modale probabilistica. Un argomento a è accettabile se, in tutti i mondi accessibili, a è difeso da tutti gli attacchi. La probabilità di accettabilità è la misura dell'insieme dei mondi in cui questa condizione è soddisfatta. Questo è un caso particolare della semantica modale probabilistica del Volume II.

In sintesi, l'argomentazione probabilistica è un laboratorio perfetto per la logica probabilistica: combina strutture non monotone, misure su mondi possibili, complessità $\#P$, semantica modale, teoria dei giochi e robustezza argomentativa. È un dominio in cui la probabilità non è un'aggiunta, ma la struttura stessa del ragionamento.

LABORATORIO 6 — Probabilità infinitesimali e funzioni di Popper

La probabilità infinitesimale è uno dei territori più sottili e concettualmente profondi dell'intera teoria della probabilità logica. È il punto in cui la misura classica incontra la non standard analysis, in cui la probabilità condizionata viene liberata dal vincolo della misura positiva, e in cui la logica condizionale trova una semantica che distingue tra impossibilità logica e improbabilità estrema. Le funzioni di Popper sono il dispositivo formale che permette di trattare la probabilità condizionata come primitiva, e non come derivata dalla probabilità assoluta. Questo rovesciamento concettuale è fondamentale: mentre nella probabilità classica $P(A|B)$ è definita solo se $P(B) > 0$, nelle funzioni di Popper $P(A|B)$ è definita per ogni B non contraddittorio, anche se $P(B)$ è infinitesimale o addirittura zero in senso standard.

Per capire tutto questo, bisogna partire dalla nozione di **campo di probabilità non standard**. Sia $\mathbb{R}$ un'estensione non standard dei reali, contenente infinitesimi ϵ tali che $0 < \epsilon < 1/n$ per ogni n standard. Una probabilità infinitesimale è un numero $p \in \mathbb{R}$ tale che $p > 0$ ma p è infinitesimo. In un modello non standard, si può definire una misura μ_{int} su un insieme iperfinito D^* come:

$$\mu_{\text{int}}(A) = |A|_{\text{int}} / |D^*|_{\text{int}},$$

dove $|A|_{\text{int}}$ è la cardinalità interna di A . Se $|D^*|_{\text{int}}$ è un iperintero infinito, allora $\mu_{\text{int}}(\{a\})$ è un infinitesimo. Questo è il punto di contatto con la misura di Loeb del Volume I: la parte standard di $\mu_{\text{int}}(A)$ è la misura classica $\mu_L(A)$, ma μ_{int} contiene informazione infinitesimale che μ_L perde. Le funzioni di Popper sono, in un certo senso, un tentativo di recuperare questa informazione infinitesimale come oggetto semantico di primo livello.

Una **funzione di Popper** è una funzione P definita su coppie di eventi (A, B) con $B \neq \emptyset$, che soddisfa tre condizioni fondamentali:

1. Per ogni B, la funzione $A \mapsto P(A|B)$ è una probabilità finitamente additiva: $P(A \wedge C|B) = P(A|B) + P(C|B)$ se $A \wedge C = \perp$, $P(B|B)=1$.
2. Catena condizionale: $P(A \wedge C|B) = P(A|B) \cdot P(C|A \wedge B)$, quando $A \wedge B$ non è impossibile.
3. Se $A \rightarrow C$ logicamente, allora $P(A|B) \leq P(C|B)$.

La differenza cruciale rispetto alla probabilità classica è che $P(A|B)$ è definita anche quando $P(B)=0$ in senso standard. Questo permette di trattare condizionali come “se B allora A” anche quando B è un evento di probabilità infinitesimale. In logica condizionale, questo è essenziale: molte frasi condizionali naturali parlano di eventi rari, non impossibili.

Per capire la semantica, consideriamo un modello non standard M^* con dominio D^* iperfinito e misura interna μ_{int} . Definiamo:

$$P(A|B) = \mu_{\text{int}}(A \wedge B) / \mu_{\text{int}}(B),$$

dove la divisione è fatta in ${}^*\mathbb{R}$. Se $\mu_{\text{int}}(B)$ è infinitesimale, $P(A|B)$ può essere un numero finito non standard, e la sua parte standard può essere qualsiasi valore in $[0,1]$. Questo permette di distinguere tra eventi di probabilità zero “forte” (impossibili) ed eventi di probabilità zero “debole” (infinitesimali ma possibili). La probabilità classica collassa entrambi in 0; la probabilità infinitesimale li distingue.

Un esempio chiarisce tutto. Consideriamo un dominio D^* con N^* elementi, dove N^* è un iperintero infinito. Sia A un sottoinsieme con $|A|_{\text{int}} = \sqrt{N^*}$, e B un sottoinsieme con $|B|_{\text{int}} = 1$. Allora:

$$\mu_{\text{int}}(A) = \sqrt{N^*} / N^* = 1/\sqrt{N^*} \text{ (infinitesimale)}, \mu_{\text{int}}(B) = 1 / N^* \text{ (infinitesimale)}, P(A|B) = (1/N) / (1/N) = 1.$$

Quindi A ha probabilità infinitesimale, B ha probabilità infinitesimale, ma condizionatamente a B, A è certo. Questo è impossibile nella probabilità classica, ma perfettamente naturale nella probabilità infinitesimale: B è un evento rarissimo, ma se accade, A accade sempre.

Questo tipo di comportamento è esattamente ciò che serve per modellare condizionali controfattuali e ragionamenti ipotetici: “se avessi lanciato la moneta mille volte e fosse uscita sempre testa, allora...”. L’evento “mille teste di fila” ha probabilità 2^{-1000} , praticamente zero, ma non è impossibile. La probabilità infinitesimale permette di trattarlo come evento condizionante.

La logica condizionale probabilistica usa le funzioni di Popper per interpretare formule del tipo:

$$\phi > \psi$$

che significa: “dato ϕ , ψ è quasi certo”. La semantica è:

$$M \models \phi > \psi \text{ se e solo se } P([\psi] \mid [\phi]) = 1 - \varepsilon, \text{ con } \varepsilon \text{ infinitesimale.}$$

Questo permette di distinguere tra:

- $\phi \rightarrow \psi$ (implicazione logica, ψ è vero in tutti i modelli che soddisfano ϕ),
- $\phi > \psi$ (implicazione probabilistica forte, ψ è vero in quasi tutti i modelli che soddisfano ϕ),
- $\phi \Rightarrow \psi$ (implicazione materiale, che è troppo debole per catturare il ragionamento condizionale naturale).

Un risultato fondamentale è che le funzioni di Popper forniscono una semantica completa per molte logiche condizionali, come la logica di Lewis, la logica di Stalnaker, e la logica probabilistica di Adams. In particolare, la logica di Adams, che interpreta “se ϕ allora ψ ” come “ $P(\psi|\phi)$ è molto alto”, è completa rispetto alle funzioni di Popper.

Un teorema centrale è il seguente.

Teorema (Adams). Una inferenza condizionale $\phi_1, \dots, \phi_n \models \psi$ è valida nella logica di Adams se e solo se, per ogni funzione di Popper P , $P(\psi | \phi_1 \wedge \dots \wedge \phi_n) = 1$.

La dimostrazione è un raffinamento della completezza di Gaifman: si costruisce una funzione di Popper che falsifica ψ se l’inferenza non è valida, usando un modello non standard in cui gli eventi ϕ_i hanno probabilità infinitesimale ma non zero.

Un altro risultato fondamentale è il collegamento tra funzioni di Popper e ranking functions di Spohn. Una ranking function κ assegna a ogni evento A un numero naturale $\kappa(A)$ che rappresenta il grado di “incredibilità” di A . La relazione con la probabilità infinitesimale è:

$$P(A) \approx \varepsilon^{\kappa(A)}.$$

Questo significa che eventi con ranking alto hanno probabilità infinitesimale molto piccola. La probabilità infinitesimale è quindi una rappresentazione quantitativa della plausibilità qualitativa.

Infine, c’è un collegamento profondo con la misura di Loeb. La misura di Loeb μ_L è la parte standard di μ_{int} . Le funzioni di Popper sono, in un certo senso, la parte non standard di μ_{int} . La misura di Loeb perde informazione infinitesimale; le funzioni di Popper la recuperano. Questo mostra che la probabilità infinitesimale è la semantica naturale dei quantificatori probabilistici del Volume I quando si vuole distinguere tra eventi di misura zero e eventi impossibili.

In sintesi, le funzioni di Popper sono il ponte tra la probabilità classica e la logica condizionale, tra la misura di Loeb e la non standard analysis, tra la plausibilità qualitativa e la probabilità quantitativa. Sono uno dei punti più profondi dell’intero trattato: permettono di trattare la probabilità come logica del condizionale, e il condizionale come probabilità infinitesimale.

LABORATORIO 7 — Fondamenti set-teorici della misurabilità

La probabilità, nella sua forma classica, è una misura definita su una σ -algebra di insiemi. Questa struttura è sufficiente per la maggior parte delle applicazioni, ma non per tutte le ambizioni teoriche. In particolare, se si vuole che **ogni sottoinsieme di $\mathbb{R}$ sia misurabile**, o che **ogni insieme definibile in un linguaggio logico abbia una misura**, o che **ogni funzione definibile sia integrabile**, allora la teoria standard della misura si rivela insufficiente. Il motivo è profondo: in ZFC, l’assioma della scelta implica l’esistenza di insiemi non misurabili. Questo è un fatto strutturale, non un accidente tecnico.

Il punto di partenza è il classico **insieme di Vitali**. Sia $\sim$ la relazione di equivalenza su $\mathbb{R}$ definita da:

$$x \sim y \Leftrightarrow x - y \in \mathbb{Q}.$$

Ogni classe di equivalenza è densa e numerabile. Usando l’assioma della scelta, si può scegliere un rappresentante per ogni classe, ottenendo un insieme $V \subseteq [0,1]$ tale che:

- V contiene esattamente un elemento per ogni classe modulo $\mathbb{Q}$;
- V è non misurabile rispetto alla misura di Lebesgue.

La dimostrazione è un capolavoro di eleganza e paradosso: se V fosse misurabile, allora le sue traslazioni razionali $V + q$, per $q \in \mathbb{Q} \cap [-1,1]$, sarebbero disgiunte e tutte di misura $\mu(V)$. Ma la loro unione è $[0,2]$, quindi:

$$\sum_{q \in \mathbb{Q} \cap [-1,1]} \mu(V) = \mu([0,2]) = 2.$$

Poiché $\mathbb{Q} \cap [-1,1]$ è infinito numerabile, l'unica possibilità sarebbe $\mu(V)=0$, ma allora la somma sarebbe 0, non 2. Contraddizione. Quindi V non è misurabile.

Questo risultato è devastante per chi desidera una teoria della probabilità “totale”: non esiste una misura di probabilità definita su tutti i sottoinsiemi di $\mathbb{R}$ che estenda la misura di Lebesgue. La probabilità classica è necessariamente **parziale**: definita solo su una σ -algebra, non sull'intero potere di $\mathbb{R}$.

Ma questo dipende dall'assioma della scelta. Se si rinuncia a AC, la situazione cambia radicalmente.

Nel 1970, Solovay dimostrò un risultato straordinario: **se esiste un cardinale inaccessibile**, allora esiste un modello di ZF + DC (Choice limitato) in cui **ogni sottoinsieme di $\mathbb{R}$ è Lebesgue misurabile**, ha la proprietà di Baire, e soddisfa il teorema di Fubini. Questo modello è noto come **modello di Solovay**.

Per capire la portata di questo risultato, bisogna introdurre i **cardinali inaccessibili**. Un cardinale κ è inaccessibile se:

1. κ è regolare: non è somma di meno di κ cardinali più piccoli;
2. κ è forte limite: per ogni $\lambda < \kappa$, $2^\lambda < \kappa$.

Gli inaccessibili sono “punti fissi” della gerarchia dei cardinali: troppo grandi per essere raggiunti da operazioni cardinali standard. La loro esistenza non è dimostrabile in ZFC, ma è coerente con esso.

Solovay mostrò che, assumendo l'esistenza di un κ inaccessibile, si può costruire un modello M di ZF + DC in cui:

- ogni insieme di reali è misurabile;
- ogni insieme di reali ha la proprietà di Baire;
- ogni funzione da $\mathbb{R}$ a $\mathbb{R}$ è quasi ovunque continua;
- non esistono insiemi di Vitali.

La costruzione è un capolavoro di forcing e teoria dei modelli. Si parte da un modello V con un cardinale inaccessibile κ . Si considera il forcing di Levy collapse $\text{Coll}(\omega, < \kappa)$, che rende $\kappa = \omega_1$ nel modello generico. Si mostra che, nel modello risultante $V[G]$, ogni insieme di reali definibile da parametri è misurabile. Poi si passa a un sotto-modello $M = \{ x \in V[G] \mid x \text{ è definibile da una sequenza di reali } \}$, che soddisfa ZF + DC e in cui ogni insieme di reali è misurabile.

Il punto concettuale è che la misurabilità universale richiede **grandi cardinali**. Non è un fatto elementare: è un fatto che vive nella parte alta della gerarchia set-teorica.

Questo ha implicazioni profonde per la logica probabilistica. Se si vuole che **ogni formula definisca un insieme misurabile**, allora bisogna assumere ipotesi forti. In ZFC, non si può garantire che ogni insieme definibile in un linguaggio logico sia misurabile. Questo limita la possibilità di definire una probabilità logica totale.

Per esempio, se si vuole definire una misura μ su tutte le formule di $L_{\omega_1\omega}$, bisogna assicurarsi che ogni insieme definito da una formula sia misurabile. Ma in ZFC, esistono formule che definiscono insiemi non misurabili. Quindi la probabilità logica non può essere totale.

Il modello di Solovay mostra che, se si rinuncia a AC e si assume un cardinale inaccessibile, allora si può avere una probabilità totale. Ma questo è un prezzo alto: si rinuncia alla scelta piena, e si assume un grande cardinale.

Un altro collegamento profondo è con la **determinacy**. L'assioma AD (Axiom of Determinacy) afferma che ogni gioco infinito a due giocatori su $\mathbb{N}$ è determinato: uno dei due ha una strategia vincente. AD implica che ogni insieme di reali è misurabile, ha la proprietà di Baire, e soddisfa il teorema di Fubini. Ma AD è incompatibile con AC. Tuttavia, AD è coerente con $ZF + DC +$ “esistono grandi cardinali”.

Questo mostra che la misurabilità universale è legata alla determinacy, non alla scelta. La probabilità totale è possibile solo in un universo deterministico, non in un universo con scelta arbitraria.

Infine, c'è un collegamento con la teoria della misura non standard. La misura di Loeb permette di definire una misura σ -additiva su insiemi interni di un modello non standard. Ma la misura di Loeb non è definita su tutti i sottoinsiemi di D^* : solo su quelli misurabili alla Carathéodory. Anche qui, la misurabilità totale è impossibile senza ipotesi forti.

In sintesi, il Laboratorio 7 mostra che la probabilità totale — una misura definita su tutti gli insiemi definibili — è possibile solo in universi set-teorici speciali, che richiedono grandi cardinali o determinacy. In ZFC, la probabilità è necessariamente parziale. Questo è un limite strutturale della logica probabilistica: non si può avere una probabilità definita per ogni formula senza assumere ipotesi forti.

LABORATORIO 8 — Apprendimento probabilistico (PAC, VC-dimension)

L'apprendimento probabilistico è il punto in cui la logica, la probabilità e la computazione si incontrano in modo più profondo. È il luogo in cui la probabilità non è più soltanto una misura su modelli, come nel Volume I, né una semantica operativa di programmi, come nel Volume II, ma diventa una **condizione epistemica**: un vincolo su ciò che un agente può imparare da dati finiti, sotto incertezza, con risorse limitate.

Il quadro teorico che governa tutto questo è il modello **PAC** (Probably Approximately Correct), introdotto da Valiant nel 1984. L'idea è sorprendentemente semplice e allo stesso tempo concettualmente rivoluzionaria: un concetto è apprendibile se, con alta probabilità, un algoritmo può trovare un'ipotesi che approssima bene il concetto vero, usando un numero di esempi che cresce polinomialmente nei parametri di precisione.

Per formalizzare tutto questo, si parte da un insieme X di possibili esempi (lo spazio degli input), e da una classe C di concetti, cioè di sottoinsiemi di X . Ogni concetto $c \in C$ è una funzione $c : X \rightarrow$

$\{0,1\}$. L'apprendimento avviene rispetto a una distribuzione D su X , sconosciuta all'algoritmo. L'algoritmo osserva m esempi i.i.d. $(x_1, y_1), \dots, (x_m, y_m)$, dove $y_i = c(x_i)$. L'obiettivo è produrre un'ipotesi $h : X \rightarrow \{0,1\}$ tale che l'errore generalizzato

$$\text{err}_D(h, c) = \Pr_{x \sim D} [h(x) \neq c(x)]$$

sia piccolo.

La definizione formale è:

Un concetto c è PAC-learnable da una classe C se esiste un algoritmo A e una funzione polinomiale $m(\epsilon, \delta)$ tali che, per ogni distribuzione D , per ogni $\epsilon, \delta > 0$, se A riceve $m \geq m(\epsilon, \delta)$ esempi da D etichettati da c , allora con probabilità $\geq 1 - \delta$ si ha $\text{err}_D(h, c) \leq \epsilon$.

Questa definizione è una forma di logica probabilistica: l'enunciato “ h approssima c ” è vero con probabilità $\geq 1 - \delta$, e l'errore è $\leq \epsilon$. La probabilità è sulla scelta del campione, non sulla verità del concetto. La logica PAC è quindi una logica della **generalizzazione**: un'ipotesi è buona se è quasi sempre vera rispetto alla distribuzione.

Il risultato fondamentale della teoria PAC è il teorema di Blumer–Ehrenfeucht–Haussler–Warmuth, che stabilisce che la PAC-learnabilità è equivalente alla finitezza della **VC-dimension**.

La VC-dimension (Vapnik–Chervonenkis) è una misura della complessità combinatoria di una classe di concetti. Formalmente, un insieme $S \subseteq X$ è detto **frammentato** da C se, per ogni funzione $f : S \rightarrow \{0,1\}$, esiste un concetto $c \in C$ tale che $c(x) = f(x)$ per ogni $x \in S$. La VC-dimension di C è il massimo k tale che esiste un insieme S di cardinalità k frammentato da C . Se non esiste tale massimo, la VC-dimension è infinita.

La VC-dimension è una misura di espressività logica: una classe C con VC-dimension infinita può rappresentare qualsiasi comportamento su insiemi arbitrariamente grandi, e quindi non può essere appresa da dati finiti. Una classe con VC-dimension finita ha una struttura combinatoria limitata, e quindi può essere appresa.

Il teorema fondamentale è:

Una classe C è PAC-learnable se e solo se $\text{VCdim}(C) < \infty$.

La dimostrazione è un capolavoro di teoria della misura e combinatoria. La direzione “VC finita $\Rightarrow$ PAC-learnabile” si basa su un lemma di uniform convergence: se $\text{VCdim}(C) = d$, allora, con $m = O((d \log(1/\epsilon) + \log(1/\delta))/\epsilon^2)$ esempi, con probabilità $\geq 1 - \delta$, per ogni $h \in C$ si ha:

$$|\text{err}_D(h, c) - \text{err}_S(h, c)| \leq \epsilon,$$

dove err_S è l'errore empirico sul campione. Questo significa che il campione è rappresentativo dell'intera distribuzione per tutte le ipotesi in C . L'algoritmo A può quindi scegliere l'ipotesi che minimizza l'errore empirico, e questa ipotesi avrà errore generalizzato $\leq \epsilon$.

La direzione opposta “VC infinita $\Rightarrow$ non PAC-learnabile” si basa su un argomento di diagonalizzazione: se $\text{VCdim}(C) = \infty$, allora per ogni m esiste un insieme S di cardinalità m frammentato da C . Si costruisce una distribuzione D concentrata su S , e si mostra che nessun algoritmo può generalizzare da m esempi, perché ogni assegnamento possibile su S è realizzabile da

un concetto in C . L'algoritmo non può distinguere tra concetti diversi che coincidono sul campione ma divergono altrove.

Questo teorema è uno dei risultati più profondi dell'intera teoria dell'apprendimento: stabilisce che la probabilità di generalizzazione è controllata dalla struttura combinatoria della classe di concetti, non dalla distribuzione dei dati. È un risultato di natura logica: la VC-dimension è una proprietà sintattica della classe C , mentre la PAC-learnabilità è una proprietà semantica rispetto a tutte le distribuzioni.

Il collegamento con la logica del primo ordine è immediato. Consideriamo una classe di concetti definibili da formule $FO(x)$ in un certo frammento della logica del primo ordine. La VC-dimension di questa classe dipende dalla profondità quantificazionale, dal numero di predicati, dall'arietà, e dalla struttura del dominio. Risultati di tipo NIP (non independence property) mostrano che frammenti di FO con restrizioni sulla profondità quantificazionale hanno VC-dimension finita. Questo significa che tali frammenti sono PAC-learnabili.

Per esempio, se X è un insieme finito e C è la classe di concetti definibili da formule FO con una sola quantificazione esistenziale, allora $VCdim(C)$ è finita. Questo significa che le formule $\exists y R(x,y)$ sono apprendibili da dati finiti. Se invece si permette la quantificazione arbitraria, la VC-dimension diventa infinita, e FO pieno non è PAC-learnabile.

Questo collega la teoria dell'apprendimento alla teoria dei modelli: la PAC-learnabilità è equivalente alla stabilità logica. Una teoria stabile ha VC-dimension finita; una teoria instabile ha VC-dimension infinita. Questo è un risultato profondo della teoria dei modelli di Shelah: la stabilità è una condizione combinatoria che controlla la complessità delle formule.

Un altro collegamento è con la teoria della misura. La PAC-learnabilità richiede che la classe C abbia una **uniform convergence**: la misura empirica converge uniformemente alla misura vera. Questo è un caso particolare del teorema di Glivenko–Cantelli, che afferma che la classe C è Glivenko–Cantelli se e solo se $VCdim(C)$ è finita. Questo è un risultato di teoria della misura: la VC-dimension controlla la dimensione metrica della classe di insiemi.

Infine, c'è un collegamento con la logica probabilistica del Volume I. La probabilità PAC è una probabilità su campioni, non su modelli. Ma la struttura è la stessa: si definisce una misura su uno spazio di mondi (i campioni), e si valuta la probabilità che una formula (l'errore) sia piccola. La differenza è che la probabilità PAC è una probabilità epistemica: misura ciò che un agente può imparare, non ciò che è vero nel mondo.

In sintesi, il Laboratorio 8 mostra che l'apprendimento probabilistico è una logica della generalizzazione: una teoria che collega la struttura combinatoria delle formule alla probabilità di apprendere da dati finiti. La VC-dimension è la misura logica della complessità; la PAC-learnabilità è la misura probabilistica della generalizzazione; la teoria dei modelli fornisce il ponte concettuale tra le due.

LABORATORIO 9 — Categorie di Markov e monade di Giry

La probabilità, quando viene osservata da una distanza sufficiente, rivela una struttura sorprendentemente semplice e al tempo stesso profondissima: ogni trasformazione probabilistica è una funzione che, invece di restituire un valore, restituisce una **misura**. Ogni composizione di trasformazioni probabilistiche è un'integrazione. Ogni sistema probabilistico è un oggetto in una categoria in cui i morfismi sono kernel di Markov. Questa prospettiva, che può sembrare astratta, è

in realtà la più naturale possibile: è la forma matematica della nozione intuitiva di “processo stocastico”.

Il cuore di tutto è la **monade di Giry**, introdotta da Michèle Giry nel 1982. La monade di Giry è un funtore G che manda uno spazio misurabile (X, Σ_X) nello spazio $G(X)$ delle misure di probabilità su X , dotato della σ -algebra generata dalle mappe $v \mapsto v(A)$ per $A \in \Sigma_X$. Ogni misura $v \in G(X)$ è un punto di $G(X)$. Ogni funzione misurabile $f : X \rightarrow Y$ induce una funzione $G(f) : G(X) \rightarrow G(Y)$ definita da:

$$G(f)(v)(B) = v(f^{-1}(B)).$$

Questa è l'immagine diretta della misura v attraverso f . La monade di Giry ha due trasformazioni naturali:

1. **unità** $\eta_X : X \rightarrow G(X)$, definita da $\eta_X(x) = \delta_x$, la delta di Dirac su x ;
2. **moltiplicazione** $\mu_X : G(G(X)) \rightarrow G(X)$, definita da $\mu_X(\Pi)(A) = \int \{v \in G(X)\} v(A) d\Pi(v)$.

Questa formula è il cuore della semantica probabilistica: una misura di misure si “appiattisce” integrando le misure interne rispetto alla misura esterna. È la formalizzazione categoriale dell'idea che “una distribuzione di distribuzioni è una distribuzione”.

La monade di Giry soddisfa gli assiomi monadici:

- $\mu_X \circ G(\eta_X) = id_{\{G(X)\}}$,
- $\mu_X \circ \eta_{\{G(X)\}} = id_{\{G(X)\}}$,
- $\mu_X \circ G(\mu_X) = \mu_X \circ \mu_{\{G(X)\}}$.

Questi assiomi sono la versione probabilistica delle leggi di associatività e unità della composizione di programmi: comporre due trasformazioni probabilistiche è un'integrazione, e l'integrazione è associativa.

La categoria di Kleisli della monade di Giry, denotata $Kl(G)$, ha come oggetti gli spazi misurabili e come morfismi le funzioni $f : X \rightarrow G(Y)$. Un tale f è un **kernel di Markov**: per ogni $x \in X$, $f(x, \cdot)$ è una misura di probabilità su Y . La composizione Kleisli è:

$$(g \circ f)(x) = \mu_Y(G(g)(f(x))),$$

che si esplicita come:

$$(g \circ f)(x)(C) = \int \{y \in Y\} g(y)(C) d f(x)(y).$$

Questa formula è la definizione stessa di composizione di kernel di Markov: si integra la distribuzione generata da f con la distribuzione generata da g . È la forma astratta della legge delle probabilità totali.

Questa struttura è sorprendentemente potente. Ogni processo stocastico discreto è un morfismo in $Kl(G)$. Ogni catena di Markov è un endomorfismo $X \rightarrow G(X)$. Ogni linguaggio di programmazione probabilistico è un linguaggio per costruire morfismi in $Kl(G)$. Ogni logica probabilistica è una logica per esprimere proprietà di tali morfismi.

Ma la prospettiva categoriale va oltre la monade di Giry. Negli ultimi anni è emersa la nozione di **categoria di Markov**, introdotta da Tobias Fritz e altri, che generalizza la struttura di $Kl(G)$ e

permette di trattare la probabilità in modo puramente diagrammatico, senza fare riferimento esplicito a spazi misurabili.

Una categoria di Markov è una categoria monoidale simmetrica in cui ogni oggetto X ha una struttura di comonoide (copy , delete):

- $\text{copy}_X : X \rightarrow X \otimes X$,
- $\text{delete}_X : X \rightarrow I$,

che soddisfano le leggi di coassociatività e counità. Queste mappe modellano la duplicazione e la marginalizzazione: copy duplica un valore, delete lo dimentica. In una categoria di Markov, i morfismi $f : X \rightarrow Y$ sono interpretati come trasformazioni probabilistiche: $f(x)$ è una distribuzione su Y . La composizione è definita tramite integrazione, come nella monade di Girmon.

La potenza di questa struttura sta nel fatto che molte proprietà probabilistiche possono essere espresse come **equazioni diagrammatiche**. Per esempio, l'indipendenza di due variabili X e Y può essere espressa come:

$\text{copy}_X \otimes \text{copy}_Y$ seguito da $f_X \otimes f_Y$

oppure come la fattorizzazione di un morfismo $X \otimes Y \rightarrow Z$ in due morfismi indipendenti.

Un risultato fondamentale è che la categoria di Kleisli della monade di Girmon è una categoria di Markov. Questo significa che la probabilità classica è un caso particolare della teoria categoriale della probabilità.

Ma la prospettiva categoriale permette di andare oltre la probabilità classica. Per esempio, si possono definire categorie di Markov in cui i morfismi sono trasformazioni **quantistiche**, e la composizione è la composizione di canali quantistici. In questo caso, la teoria categoriale unifica la probabilità classica e la probabilità quantistica.

Un altro risultato profondo è che i teoremi di rappresentazione di de Finetti e Aldous–Hoover possono essere formulati in termini categoriali. In una categoria di Markov, un morfismo $X^{\mathbb{N}} \rightarrow Y$ è scambiabile se è invariato sotto l'azione del gruppo delle permutazioni. Il teorema di de Finetti afferma che ogni morfismo scambiabile fattorizza attraverso un oggetto di parametri latenti Θ :

$$X^{\mathbb{N}} \rightarrow \Theta \rightarrow Y^{\mathbb{N}}.$$

Questa è la forma categoriale della rappresentazione di de Finetti: una sequenza scambiabile è una miscela di sequenze i.i.d. La categoria di Markov fornisce il linguaggio naturale per esprimere questa fattorizzazione.

La monade di Girmon è quindi il cuore semantico della probabilità, e le categorie di Markov sono il linguaggio strutturale che permette di esprimere la probabilità in modo astratto, diagrammatico, composizionale. La logica probabilistica, in questa prospettiva, è una logica degli invarianti categoriali: un enunciato è vero se è preservato da tutti i morfismi probabilistici.

In sintesi, il Laboratorio 9 mostra che la probabilità non è solo una misura su insiemi, ma una struttura categoriale: una teoria della composizione di trasformazioni stocastiche. La monade di Girmon è la forma analitica di questa struttura; le categorie di Markov sono la forma diagrammatica; la logica probabilistica è la forma linguistica. Tutte e tre sono facce della stessa realtà matematica.

LABORATORIO 10 — Logiche modali e temporali probabilistiche

Quando si passa dalla logica statica alla logica dinamica, la probabilità assume un ruolo completamente nuovo. Non è più soltanto una misura su insiemi di modelli, come nel Volume I, né una semantica di programmi probabilistici, come nel Volume II, ma diventa una **misura sui cammini**: una misura su sequenze infinite di stati generate da un sistema dinamico. È qui che nascono le logiche modali probabilistiche e le logiche temporali probabilistiche, come PCTL, PCTL*, probabilistic μ -calculus, probabilistic CTL, probabilistic LTL, e molte altre varianti.

Il punto di partenza è un modello probabilistico del comportamento dinamico: una **catena di Markov** o un **processo decisionale di Markov** (MDP). Una catena di Markov è una tripla:

$$M = (S, P, L),$$

dove S è un insieme di stati, $P : S \times S \rightarrow [0,1]$ è la matrice di transizione, e $L : S \rightarrow 2^{\{AP\}}$ è una funzione di etichettatura che assegna a ogni stato gli atomi proposizionali veri in quello stato. La condizione fondamentale è che, per ogni $s \in S$,

$$\sum_{t \in S} P(s,t) = 1.$$

Una catena di Markov genera una misura su S^ω , lo spazio dei cammini infiniti. La misura cilindrica è definita da:

$$\Pr_s(s_0, s_1, \dots, s_n) = \prod_{i=0}^{n-1} P(s_i, s_{i+1}),$$

con $s_0 = s$. Questa misura si estende univocamente a una misura su S^ω tramite il teorema di Carathéodory. Ogni formula temporale probabilistica è interpretata rispetto a questa misura.

La logica temporale probabilistica più studiata è **PCTL** (Probabilistic Computation Tree Logic). Le formule di stato sono:

$$\varphi ::= \text{true} \mid a \mid \neg\varphi \mid \varphi \wedge \varphi \mid P_{\geq p}[\psi],$$

dove a è un atomo proposizionale e ψ è una formula di cammino:

$$\psi ::= X\varphi \mid \varphi \cup \varphi \mid \varphi U^{\leq k} \varphi.$$

La semantica è definita come segue. Una formula $P_{\geq p}[\psi]$ è vera in uno stato s se la misura dei cammini che partono da s e soddisfano ψ è $\geq p$. Formalmente:

$$M, s \models P_{\geq p}[\psi] \Leftrightarrow \Pr_s(\{ \pi \in S^\omega \mid \pi \models \psi \}) \geq p.$$

Questa definizione è il cuore della logica temporale probabilistica: la probabilità è una misura su cammini, non su stati. La logica modale probabilistica del Laboratorio 3 era una logica su transizioni; qui la probabilità è una misura su sequenze infinite.

La formula $X\varphi$ (“next φ ”) è vera su un cammino $\pi = s_0, s_1, \dots$ se $s_1 \models \varphi$. La formula $\varphi \cup \psi$ (“ φ until ψ ”) è vera se esiste un indice i tale che $s_i \models \psi$ e, per ogni $j < i$, $s_j \models \varphi$. La formula $\varphi U^{\leq k} \psi$ è la versione con bound temporale.

Il problema fondamentale è il **model checking probabilistico**: dato un modello M e una formula φ , decidere se $M, s \models \varphi$. Per PCTL, il model checking è in tempo polinomiale. La ragione è che la probabilità di soddisfare una formula until può essere calcolata risolvendo un sistema lineare. Per esempio, per la formula $\varphi \cup \psi$, si definisce una funzione $x : S \rightarrow [0,1]$ tale che:

$$x(s) = 1 \text{ se } s \models \psi, \quad x(s) = 0 \text{ se } s \not\models \varphi \wedge s \not\models \psi, \quad x(s) = \sum_{t \in S} P(s,t) x(t) \text{ se } s \models \varphi \wedge s \not\models \psi.$$

Questo è un sistema lineare $x = A x + b$, che ha una soluzione unica perché A è una matrice stocastica sub-unitaria. La soluzione $x(s)$ è la probabilità che un cammino partendo da s soddisfi $\varphi \cup \psi$. Questa è una forma di equazione di Bellman, e la sua soluzione è un caso particolare del teorema di Banach sul punto fisso.

Un risultato fondamentale è che PCTL è **bisimulation-invariante**: se due stati s e t sono probabilisticamente bisimili, allora soddisfano le stesse formule PCTL. La probabilistic bisimulation è definita come nel Laboratorio 3: due stati sono bisimili se hanno la stessa etichettatura e le loro distribuzioni di transizione sono equivalenti modulo la relazione. Il teorema di Hennessy–Milner probabilistico afferma che la bisimulazione è esattamente la relazione di equivalenza per PCTL.

Questo risultato è cruciale per la verifica: se si può ridurre un modello grande a un modello più piccolo tramite bisimulazione, allora si può fare model checking sul modello ridotto.

Ma PCTL non è l'unica logica temporale probabilistica. Esiste anche **PCTL**, *che permette formule di cammino più complesse, come combinazioni booleane di until e next. La semantica è la stessa, ma il model checking è molto più difficile: PCTL ha complessità PSPACE-complete.*

Esiste poi il **probabilistic μ -calculus**, che estende il μ -calculus classico con operatori probabilistici. Le formule sono costruite da:

$$\varphi ::= a \mid X \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid P_{\leq p}(\varphi) \mid \mu X. \varphi \mid \nu X. \varphi.$$

Gli operatori μ e ν sono punti fissi minimo e massimo. La semantica è definita tramite funzioni monotone su $[0,1]^S$. Per esempio, la formula $\mu X. \varphi(X)$ è interpretata come il punto fisso minimo della funzione $F(x) = \llbracket \varphi \rrbracket(x)$. Questo è un caso particolare del teorema di Tarski sui punti fissi. Il probabilistic μ -calculus è estremamente espressivo: può esprimere tutte le proprietà ω -regolari probabilistiche.

Un risultato profondo è che il probabilistic μ -calculus è **bisimulation-invariante e equivalente a PCTL*** in potere espressivo. La dimostrazione usa tecniche di teoria dei giochi: ogni formula del μ -calculus può essere interpretata come un gioco di parità probabilistico, e la probabilità di vittoria è la semantica della formula.

Un altro risultato fondamentale è la **legge zero-uno per PCTL** su modelli random. Se si considera una famiglia di grafi random $G(n,p)$, allora per ogni formula PCTL φ , la probabilità che $G(n,p) \models \varphi$ tende a 0 o 1 quando $n \rightarrow \infty$. Questo è un caso particolare delle leggi zero-uno del Volume I, applicate a logiche temporali. La dimostrazione usa tecniche di convergenza dei grafi e di teoria dei limiti.

Infine, c'è un collegamento profondo con la semantica dei programmi probabilistici del Laboratorio 2. Un programma probabilistico può essere visto come un kernel di Markov, e la sua esecuzione come un cammino in una catena di Markov. Le proprietà del programma possono essere espresse in PCTL: la probabilità che il programma termini, la probabilità che violi un invariante, la probabilità che

raggiunga uno stato di errore. La logica temporale probabilistica è quindi la logica naturale per la verifica quantitativa dei programmi probabilistici.

In sintesi, il Laboratorio 10 mostra che la probabilità, quando applicata ai sistemi dinamici, diventa una misura sui cammini, e la logica probabilistica diventa una logica temporale. PCTL, PCTL*, probabilistic μ -calculus sono i linguaggi naturali per esprimere proprietà quantitative dei sistemi probabilistici. La bisimulazione probabilistica è la relazione di equivalenza naturale. La teoria dei punti fissi, la teoria della misura e la teoria dei giochi si intrecciano in modo profondo.

LABORATORIO 11 — Logiche infinitarie con probabilità ($L_{\omega_1\omega, P}$, leggi zero–uno)

Le logiche infinitarie sono uno dei territori più affascinanti della logica matematica: permettono congiunzioni e disgiunzioni numerabili, pur mantenendo la lunghezza finita dei blocchi di quantificatori. La logica $L_{\omega_1\omega}$ è il prototipo: consente formule come $\bigwedge_{n \in \mathbb{N}} \varphi_n$, $\bigvee_{n \in \mathbb{N}} \psi_n$, ma non consente quantificazioni infinite. È una logica infinitamente più espressiva della logica del primo ordine, ma ancora abbastanza controllata da permettere teoremi di compattezza parziale, teoremi di omogeneità, e soprattutto **leggi zero–uno** in modelli random.

Quando si aggiungono **quantificatori probabilistici**, come $(P x \geq r) \varphi(x)$, si ottiene una logica infinitaria probabilistica, che chiameremo $L_{\omega_1\omega, P}$. Questa logica è in grado di esprimere proprietà globali dei modelli che nessuna logica finitaria può esprimere: densità, limiti, convergenze, frequenze asintotiche, comportamenti globali di strutture infinite.

Per capire la semantica, bisogna partire da un modello M con dominio D , e da una misura μ su D , tipicamente una misura di Loeb o una misura counting normalizzata su un dominio finito. La semantica del quantificatore probabilistico è:

$$M \models (P x \geq r) \varphi(x) \text{ se e solo se } \mu(\{a \in D \mid M \models \varphi(a)\}) \geq r.$$

Questa definizione è identica a quella del Volume I, ma ora φ può essere una formula infinitaria. Questo permette di esprimere proprietà come:

$$(P x \geq 1) (\bigwedge_{n=1}^{\infty} \varphi_n(x)),$$

che significa che **quasi tutti gli elementi del dominio soddisfano tutte le φ_n** , cioè soddisfano un'infinità di condizioni simultaneamente. Questo è impossibile da esprimere in FO, e anche in molte logiche modali probabilistiche.

La potenza espressiva di $L_{\omega_1\omega, P}$ è enorme: può esprimere limiti di sequenze di proprietà, densità di sottostrutture, convergenze di frequenze, e persino proprietà di tipo ergodico. Ma questa potenza ha un prezzo: la logica non è compatta, e la sua teoria dei modelli è molto più complessa di quella di FO.

Il fenomeno più sorprendente è la **legge zero–uno**. In molte classi di modelli random, come i grafi $G(n, p)$, per ogni formula φ di $L_{\omega_1\omega, P}$, la probabilità che $G(n, p) \models \varphi$ tende a 0 o 1 quando $n \rightarrow \infty$. Questo significa che la logica infinitaria probabilistica non riesce a distinguere tra proprietà che hanno probabilità intermedia: ogni proprietà o è quasi sempre vera, o è quasi sempre falsa.

Per capire questo fenomeno, consideriamo il caso classico dei grafi random $G(n,p)$, con p costante in $(0,1)$. La legge zero–uno di Glebskii–Kogan–Liogon’kii–Talanov e, indipendentemente, di Fagin, afferma che per ogni formula FO φ ,

$$\lim_{n \rightarrow \infty} \Pr[G(n,p) \models \varphi] \in \{0,1\}.$$

Questo è un risultato profondo: la logica del primo ordine non può esprimere proprietà che hanno probabilità limite intermedia. Per esempio, la proprietà “il grafo ha un numero pari di vertici” non è esprimibile in FO, perché ha probabilità $1/2$.

Quando si passa a $L\{\omega_1\omega\}$, la situazione cambia: la logica può esprimere proprietà molto più complesse, come “il grafo ha un numero infinito di cicli di lunghezza crescente”. Ma la legge zero–uno continua a valere per molti frammenti di $L\{\omega_1\omega\}$. Il motivo è che, nei grafi random, la struttura locale è altamente omogenea, e la struttura globale è deterministica in senso probabilistico.

Quando si aggiungono quantificatori probabilistici, la situazione diventa ancora più interessante. La logica $L\{\omega_1\omega, P\}$ può esprimere proprietà come:

$$(P x \geq p) \varphi(x),$$

e ciò significa che la densità degli elementi che soddisfano φ è $\geq p$. Nei grafi random, la densità di molte proprietà converge a un valore deterministico, e quindi la formula ha probabilità limite 0 o 1.

Per esempio, la proprietà “la densità dei vertici di grado $\geq k$ è $\geq r$ ” ha probabilità limite 1 se r è minore della densità limite, e 0 se r è maggiore. Questo è un caso particolare della legge dei grandi numeri applicata ai grafi random.

Il risultato generale è il seguente.

Teorema (Legge zero–uno per $L\{\omega_1\omega, P\}$ su $G(n,p)$). Per ogni formula φ di $L\{\omega_1\omega, P\}$, $\lim_{n \rightarrow \infty} \Pr[G(n,p) \models \varphi] \in \{0,1\}$.

La dimostrazione è un’estensione della dimostrazione classica per FO. Si basa su tre ingredienti:

1. **Omogeneità locale:** ogni grafo $G(n,p)$ è quasi sicuramente omogeneo su scale finite.
2. **Concentrazione della misura:** la densità di ogni proprietà locale converge quasi sicuramente a un valore deterministico.
3. **Induzione strutturale** sulle formule infinitarie: le congiunzioni numerabili e i quantificatori probabilistici preservano la convergenza deterministica.

Il punto cruciale è che, in $G(n,p)$, la misura counting normalizzata converge a una misura limite deterministica, e quindi ogni quantificatore probabilistico ha un valore limite deterministico. Le congiunzioni numerabili non introducono oscillazioni, perché ogni φ_n ha un limite deterministico, e quindi la congiunzione ha un limite deterministico.

Questo risultato mostra che la logica infinitaria probabilistica, pur essendo molto più espressiva di FO, non riesce a catturare proprietà che hanno probabilità limite intermedia. La probabilità limite è sempre 0 o 1.

Un altro fenomeno interessante è la **completezza asintotica**. In molti modelli random, esiste una teoria $T\{\lim\}$ tale che, per ogni formula φ , $\lim_{n \rightarrow \infty} \Pr[M_n \models \varphi] = 1$ se e solo se $T\{\lim\} \models \varphi$.

Questa teoria è la teoria limite del modello random. Per $G(n,p)$, $T\{\lim\}$ è la teoria dei grafi aleatori di Rado. Questa teoria è completa, ω -categorica, e ha eliminazione dei quantificatori. La logica $L\{\omega_1\omega, P\}$ può esprimere molte proprietà di $T\{\lim\}$, ma non può distinguere tra modelli che differiscono su proprietà di probabilità intermedia.

Infine, c'è un collegamento profondo con la teoria dei limiti dei grafi di Lovász. Ogni sequenza di grafi densi ha un limite in uno spazio compatto di oggetti chiamati graphons. La logica $L\{\omega_1\omega, P\}$ è, in un certo senso, la logica naturale per descrivere proprietà dei graphons: i quantificatori probabilistici corrispondono alle integrali sui graphons, e le congiunzioni numerabili corrispondono a proprietà di convergenza.

In sintesi, il Laboratorio 11 mostra che la logica infinitaria probabilistica è una logica dei limiti: una logica che descrive proprietà globali di strutture infinite, densità, convergenze, comportamenti asintotici. Le leggi zero–uno mostrano che, in molti modelli random, questa logica è asintoticamente deterministica: ogni proprietà ha probabilità limite 0 o 1. La teoria dei limiti dei grafi fornisce la semantica naturale per questa logica.

LABORATORIO 12 — Rappresentazioni scambiabili (Aldous–Hoover, Hoover definibilità)

La scambiabilità è uno dei concetti più profondi della probabilità moderna. È l'idea che l'ordine degli indici non conti: ciò che importa è la struttura relazionale tra gli oggetti, non la loro etichetta. Questa idea, che in apparenza è solo un principio di simmetria, si rivela essere la chiave per comprendere la struttura delle misure su sequenze, matrici, ipergrafi, strutture relazionali infinite, e persino modelli di logica del primo ordine.

Il punto di partenza è il teorema di de Finetti. Una sequenza di variabili casuali $\{X_n\}$ è scambiabile se la sua distribuzione è invariante sotto permutazioni finite:

$$(X_1, X_2, \dots) \equiv (X_{\pi(1)}, X_{\pi(2)}, \dots)$$

per ogni permutazione finita π . Il teorema di de Finetti afferma che ogni sequenza scambiabile è una miscela di sequenze i.i.d.:

$$\Pr(X_1=x_1, \dots, X_n=x_n) = \int \prod_{i=1}^n \mu(x_i) d\Pi(\mu).$$

La misura Π è la distribuzione dei parametri latenti. Questo teorema è un caso particolare di un principio molto più generale: **ogni struttura scambiabile è una miscela di strutture indipendenti condizionate a parametri latenti.**

Quando si passa da sequenze a matrici, il teorema di de Finetti non basta più. Una matrice $\{X_{ij}\}$ è scambiabile se la sua distribuzione è invariante sotto permutazioni simultanee delle righe e delle colonne:

$$(X_{ij}) \equiv (X_{\pi(i)\pi(j)}).$$

Il teorema di Aldous–Hoover afferma che ogni matrice scambiabile può essere rappresentata come:

$$X_{ij} = f(U, U_i, U_j, U_{ij}),$$

dove:

- U è un parametro globale,
- U_i è un parametro associato alla riga i ,
- U_j è un parametro associato alla colonna j ,
- $U_{\{ij\}}$ è un rumore indipendente,
- f è una funzione misurabile.

Questa rappresentazione è sorprendente: ogni matrice scambiabile è generata da una funzione f che combina un parametro globale, parametri individuali e rumore indipendente. È la versione bidimensionale del teorema di de Finetti.

Ma il teorema di Aldous–Hoover è solo l’inizio. Hoover generalizzò questa rappresentazione a strutture relazionali di arità arbitraria. Una struttura relazionale M con dominio N e relazioni $R_1, \dots, R_k$ è scambiabile se la sua distribuzione è invariante sotto permutazioni finite di N :

$$M \equiv \pi(M).$$

Il teorema di Hoover afferma che ogni struttura scambiabile può essere rappresentata come:

$$R(a_1, \dots, a_r) = f_R(U, U_{\{a_1\}}, \dots, U_{\{a_r\}}, U_{\{a_1 a_2\}}, \dots, U_{\{a_1 \dots a_r\}}),$$

dove:

- U è un parametro globale,
- U_i è un parametro associato all’elemento i ,
- $U_{\{ij\}}$ è un parametro associato alla coppia $\{i, j\}$,
- $U_{\{ijk\}}$ è un parametro associato alla tripla $\{i, j, k\}$,
- e così via fino all’arietà r .

Ogni $U_{\{i_1 \dots i_k\}}$ è una variabile casuale indipendente uniforme su $[0, 1]$, simmetrica rispetto alle permutazioni degli indici. La funzione f_R è misurabile e determina il valore della relazione R .

Questa rappresentazione è estremamente potente: ogni misura scambiabile su strutture relazionali è una miscela di modelli deterministici generati da parametri latenti. È la versione strutturale del teorema di de Finetti.

Ma il punto più profondo è la connessione con la **definibilità logica**. Una misura μ su strutture di un linguaggio L è **definibile** se, per ogni formula $\varphi(x)$, la funzione:

$$M \mapsto \mu_M(\{a \mid M \models \varphi(a)\})$$

è approssimabile da formule quantificatore-free. Questo significa che la probabilità che un elemento soddisfi φ è determinata da proprietà locali, non da proprietà globali complesse.

Il teorema di Hoover sulla definibilità afferma che una misura scambiabile è definibile se e solo se la funzione f_R nella rappresentazione di Aldous–Hoover dipende solo da parametri di livello inferiore. In altre parole, la definibilità logica corrisponde a una restrizione strutturale sulla funzione generatrice.

Per esempio, una misura è definibile se ogni relazione $R(a,b)$ dipende solo da U , U_a , U_b , e non da $U_{\{ab\}}$. Questo significa che la relazione è determinata da proprietà individuali, non da rumore di coppia. Questo è esattamente ciò che accade nei modelli di tipo “latent feature”: ogni elemento ha un vettore di caratteristiche U_a , e le relazioni sono determinate da queste caratteristiche.

Il collegamento con la logica è profondo. Una misura definibile corrisponde a una teoria stabile nel senso di Shelah. Una misura non definibile corrisponde a una teoria instabile. La rappresentazione di Aldous–Hoover è quindi una rappresentazione probabilistica della stabilità logica.

Un altro collegamento è con i **graphons**, i limiti dei grafi densi. Un graphon è una funzione $W : [0,1]^2 \rightarrow [0,1]$. Un grafo random scambiabile può essere generato come:

$$X_{\{ij\}} = \text{Bernoulli}(W(U_i, U_j)).$$

Questo è un caso particolare della rappresentazione di Aldous–Hoover. La teoria dei graphons è quindi un caso particolare della teoria delle strutture scambiabili.

Infine, c’è un collegamento con la teoria dei modelli continua. La rappresentazione di Aldous–Hoover può essere vista come una rappresentazione di modelli in logica continua, in cui le relazioni sono funzioni $[0,1]$ -valutate. La scambiabilità corrisponde all’omogeneità, e la rappresentazione corrisponde alla decomposizione di un modello in parametri latenti.

In sintesi, il Laboratorio 12 mostra che la scambiabilità è la chiave per comprendere la struttura delle misure su strutture infinite. Il teorema di Aldous–Hoover fornisce una rappresentazione universale. La definibilità di Hoover collega questa rappresentazione alla logica del primo ordine. La teoria dei graphons è un caso particolare. La teoria dei modelli continua fornisce il quadro concettuale. Tutto converge in un’unica visione: le strutture scambiabili sono miscele di modelli deterministici generati da parametri latenti.

LABORATORIO 13 — Politopi probabilistici proposizionali (Fagin–Halpern–Megiddo, Nilsson)

La probabilità proposizionale è il punto in cui la logica più elementare — quella delle proposizioni booleane — incontra la geometria convessa. È un incontro sorprendente: la semantica probabilistica delle formule proposizionali non è un oggetto combinatorio, ma un oggetto geometrico, un **politopo** in uno spazio euclideo di dimensione esponenziale. Questo politopo è la chiave per comprendere la probabilità logica proposizionale, la coerenza probabilistica, l’inferenza probabilistica, e la complessità computazionale delle logiche probabilistiche proposizionali.

Il punto di partenza è semplice. Sia Φ un insieme finito di proposizioni atomiche $p_1, \dots, p_n$. Ogni interpretazione ω è una funzione $\omega : \Phi \rightarrow \{0,1\}$. L’insieme Ω di tutte le interpretazioni ha cardinalità 2^n . Una misura di probabilità μ su Ω è un punto del semplice standard $\Delta_{\{2^n-1\}}$, cioè un vettore:

$$\mu = (\mu(\omega_1), \dots, \mu(\omega_{\{2^n\}})),$$

$$\text{con } \mu(\omega_i) \geq 0 \text{ e } \sum \mu(\omega_i) = 1.$$

Ogni formula proposizionale φ definisce un sottoinsieme $[\varphi] \subseteq \Omega$, e la probabilità di φ è:

$$\text{Pr}(\varphi) = \sum_{\{\omega \models \varphi\}} \mu(\omega).$$

Questa è la semantica di Gaifman nel caso proposizionale. Ma ciò che rende tutto interessante è che **ogni vincolo probabilistico su formule proposizionali è una disuguaglianza lineare su μ** . Per esempio:

$$\Pr(p \wedge q) \geq 0.3 \quad \Pr(p) + \Pr(q) \leq 1.2 \quad \Pr(p \rightarrow q) = 1$$

sono tutte disuguaglianze lineari su μ . Questo significa che l'insieme delle misure che soddisfano un insieme di vincoli probabilistici è un politopo convesso.

Nilsson fu il primo a osservare questo fenomeno in modo sistematico. Nel suo lavoro del 1986, mostrò che la probabilità proposizionale è un problema di programmazione lineare: dato un insieme di vincoli probabilistici, si può verificare la loro coerenza risolvendo un sistema lineare. Se il sistema ha una soluzione, i vincoli sono coerenti; se non ha soluzione, sono incoerenti.

Ma il punto più profondo è che l'insieme di tutte le misure che soddisfano i vincoli è un politopo convesso in $\mathbb{R}^{2^n}$. Questo politopo è chiamato **politopo delle probabilità proposizionali**. La sua struttura è estremamente ricca: le sue facce corrispondono a vincoli logici, i suoi vertici corrispondono a interpretazioni deterministiche, e la sua dimensione è $2^n - 1$.

Il lavoro di Fagin, Halpern e Megiddo (FHM) portò questa teoria a un livello completamente nuovo. Essi considerarono una logica probabilistica proposizionale in cui le formule sono costruite da:

- formule proposizionali φ ,
- espressioni probabilistiche come $\Pr(\varphi) \geq r$,
- combinazioni booleane di tali espressioni.

La semantica è definita rispetto a una misura μ su Ω . La domanda fondamentale è: **quando un insieme di vincoli probabilistici è soddisfacibile?** La risposta è: **quando il sistema lineare corrispondente ha una soluzione.**

Formalmente, ogni vincolo $\Pr(\varphi) \geq r$ è tradotto in:

$$\sum_{\omega \models \varphi} \mu(\omega) \geq r.$$

Ogni vincolo $\Pr(\varphi) \leq r$ è tradotto in:

$$\sum_{\omega \models \varphi} \mu(\omega) \leq r.$$

Ogni vincolo $\Pr(\varphi) = r$ è tradotto in:

$$\sum_{\omega \models \varphi} \mu(\omega) = r.$$

L'insieme dei vincoli è soddisfacibile se e solo se il sistema lineare ha una soluzione μ con $\mu \geq 0$ e $\sum \mu = 1$. Questo è un risultato di coerenza probabilistica: la probabilità proposizionale è esattamente la teoria dei politopi convessi.

Ma la cosa essenziale di FHM è che **la logica probabilistica proposizionale è decidibile in tempo polinomiale**. Il motivo è che il sistema lineare ha dimensione esponenziale, ma può essere risolto senza enumerare tutte le 2^n interpretazioni. FHM mostrarono che il problema può essere ridotto a un sistema lineare di dimensione polinomiale, usando tecniche di eliminazione e rappresentazione compatta delle formule.

Il risultato è sorprendente: la logica probabilistica proposizionale è molto più semplice della logica proposizionale classica, che è NP-completa. La probabilità “ammorbidisce” la logica: la rende convessa, e quindi computazionalmente più semplice.

Ma la storia non finisce qui. Il politopo delle probabilità proposizionali ha una struttura geometrica profonda. I suoi vertici sono le interpretazioni deterministiche. Le sue facce corrispondono a vincoli logici come:

$$\Pr(p \wedge q) \leq \Pr(p), \Pr(p \vee q) \geq \Pr(p), \Pr(p) + \Pr(\neg p) = 1.$$

Questi vincoli sono le **leggi della probabilità**. FHM mostrarono che tutte le leggi della probabilità proposizionale sono disuguaglianze lineari valide per il politopo. Questo significa che la probabilità proposizionale è una teoria geometrica: le sue leggi sono le facce del politopo.

Un risultato particolarmente elegante è che la logica probabilistica proposizionale è **completa** rispetto alla semantica dei politopi. Ogni formula probabilistica valida è una conseguenza delle leggi della probabilità. Questo è un teorema di completezza geometrica: la logica è esattamente la teoria del politopo.

Infine, c'è un collegamento profondo con la teoria della misura. Ogni misura μ su Ω è una misura atomica, e quindi la probabilità proposizionale è una teoria delle misure atomiche. Ma il politopo è la convessità di tutte le misure atomiche. Questo significa che la probabilità proposizionale è la teoria delle misure atomiche e delle loro miscele. È un caso particolare della teoria delle misure su modelli del Volume I.

In sintesi, il Laboratorio 13 mostra che la probabilità proposizionale è una teoria geometrica: una teoria dei politopi convessi. La logica probabilistica proposizionale è una teoria della coerenza lineare. Il politopo delle probabilità proposizionali è la struttura geometrica che governa la semantica. Il lavoro di Nilsson e FHM mostra che la probabilità proposizionale è sorprendentemente semplice dal punto di vista computazionale, e sorprendentemente ricca dal punto di vista geometrico.

Questa parte è, in un certo senso, una metalogica della logica probabilistica. Non si limita più a usare la misura come strumento semantico, ma la analizza come oggetto logico in sé, in diverse cornici: topos e logiche intuizioniste, probabilità quantistica, semantica a punti fissi per sistemi dinamici infiniti, teorie dell'informazione, categorie coalgebriche. Il filo rosso rimane sempre lo stesso: le nozioni chiave dei Volumi I e II (misure su modelli, quantificatori probabilistici, kernel di Markov, bisimulazione probabilistica, scambiabilità) vengono riformulate in strutture più generali, dove diventano casi particolari di principi più profondi.

Probabilità nei topos e logica interna

Nella parte iniziale di questo contributo la probabilità era una misura su una σ -algebra di insiemi; poi era una monade su spazi misurabili, e i programmi probabilistici erano kernel di Markov. Una generalizzazione naturale consiste nel chiedersi: cosa succede se la “logica di base” non è più classica, ma intuizionista? Se, al posto di Set, consideriamo un topos $\mathbb{I}$ come universo fondamentale, esiste una nozione interna di probabilità?

Un topos elementare $\mathbb{I}$ ha un oggetto dei sottoinsiemi, il classificatore di sottooggetti Ω , che interpreta la logica intuizionista: ogni sottooggetto $U \hookrightarrow X$ è classificato da una caratteristica $\chi_U :$

$X \rightarrow \Omega$, e gli elementi di Ω rappresentano verità parziali. In Set, $\Omega = \{0,1\}$; in un topos generale, Ω è un oggetto più ricco, con struttura di Heyting algebra, non necessariamente booleana.

Per parlare di probabilità interna, l'idea è: al posto di misure $\mu : \Sigma \rightarrow [0,1]$, considerare morfismi

$$v : \Omega \rightarrow [0,1],$$

dove $[0,1]$ è implementato come un oggetto interno di $\mathbb{I}$ (per esempio, *reals object*). Ma questo è troppo rozzo: v così definito non “vede” il fatto che le formule non sono semplicemente sottoinsiemi, bensì suboggetti in logica intuizionista.

La via più promettente è introdurre una nozione di “valutazione” alla Kripke–Joyal: una misura interna non assegna direttamente numeri a sottooggetti, ma è un morfismo

$$\mu : \mathbb{K}(X) \rightarrow \mathbb{I},$$

dove $\mathbb{K}(X)$ è un oggetto di sottooggetti di X , e $\mathbb{I}$ è un oggetto che rappresenta “intensità” o “gradi di credenza”. In un topos booleano classico, $\mathbb{I}$ può essere (una copia interna di) $[0,1]$; in un topos più generale, $\mathbb{I}$ può essere un oggetto con struttura di quantale commutativo, in cui l'additività è espressa come

$$\mu(U \vee V) = \mu(U) \oplus \mu(V),$$

quando $U \wedge V$ è nullo. La σ -additività diventa allora un assioma interno espresso in logica intuizionista, usando quantificazioni interne $\forall$ e $\exists$.

Questo porta a una variante della logica probabilistica: i quantificatori probabilistici ($P x \geq r$) $\varphi(x)$ non sono più interpretati in Set, ma nella logica interna del topos. In pratica, si ottiene una logica probabilistica intuizionista, in cui la probabilità non è solo un numero, ma un costrutto interno. La dualità con la monade di Girard è che, in un topos di sheaves, si può definire un “Giry interno” che manda un oggetto X al fascio delle misure sui sezioni locali di X . La probabilità diventa quindi uno stratificato locale: una misura definita “pezzetto per pezzetto”, con *gluing conditions* di tipo sheaf.

Questo cambia anche il modo di intendere le leggi zero–uno: in una logica intuizionista interna, l'asserzione “ $\mu(\varphi) = 1$ ” non è necessariamente equivalente a “ $\neg\neg\varphi$ ”, ma solo a un certo rafforzamento della sua dimostrabilità interna. La probabilità interna misura non più soltanto grandezze cardinali, ma stabilità intuitionista di asserzioni.

Questa linea porta in direzione dei **topos probabilistici**: categorie in cui ogni oggetto X porta con sé uno spazio di misure $G(X)$ e una struttura monadica compatibile con l'interno. La logica probabilistica, vista da qui, è la logica interna di un topos arricchito: fissa non solo cosa significa “vero”, ma anche cosa significa “probabile”.

Probabilità quantistica e logica quantistica

Se abbandoniamo non solo la booleanità ma persino la distributività, entriamo nella probabilità quantistica. Qui, gli “eventi” non sono più sottoinsiemi di uno spazio, ma proiezioni in uno spazio di Hilbert. In una algebra di von Neumann $\mathfrak{A}$ (un'algebra di operatori chiusa in topologia debole, con unità), l'insieme delle proiezioni $\text{Proj}(\mathfrak{A})$ forma un reticolo ortomodulare, che prende il posto della σ -algebra classica.

Una **stato** quantistico è una funzione

$$q : \frac{\mathbb{M}}{\sim} \rightarrow \mathbb{C}$$

che è lineare, positiva ($q(A^* A) \geq 0$), normalizzata ($q(1)=1$). Su $\text{Proj}(\frac{\mathbb{M}}{\sim})$, la restrizione di q assegna una probabilità a ogni proiezione P : $q(P)$ è la probabilità che l'evento rappresentato da P si verifichi. Ma il reticolo $\text{Proj}(\frac{\mathbb{M}}{\sim})$ non è distributivo: in generale,

$$P \wedge (Q \vee R) \neq (P \wedge Q) \vee (P \wedge R).$$

Questo rompe le leggi del calcolo delle probabilità classico.

Il teorema di Gleason mostra che, in dimensione ≥ 3 , ogni funzione

$$\mu : \text{Proj}(\mathcal{H}) \rightarrow [0,1]$$

che è finitamente additiva su famiglie ortogonali (cioè se $P_i P_j = 0$ per $i \neq j$, e $\sum P_i = 1$, allora $\sum \mu(P_i)=1$) è indotta da uno stato q su $B(\mathcal{H})$: esiste un operatore di traccia positiva ρ con $\text{tr}(\rho)=1$ tale che $\mu(P)=\text{tr}(\rho P)$. Questo significa che la misura quantistica sugli eventi è equivalentemente descritta da stati densità.

La logica quantistica nasce dal tentativo di interpretare $\text{Proj}(\mathcal{H})$ come algebra di proposizioni, con $\wedge$, $\vee$, $\perp$ (ortocomplemento) come congiunzione, disgiunzione, negazione. La probabilità $q(P)$ diventa allora la probabilità che la proposizione P sia vera. Le difficoltà sono enormi: la mancanza di distributività implica che “ $P \wedge (Q \vee R)$ ” non si comporta più come ci si aspetta da una logica classica. Ciò che viene salvato è la struttura ortomodulare: un compromesso tra logica classica e struttura empirica quantistica.

Dal nostro punto di vista, il fatto decisivo è che qui la **misura su modelli** di cui *supra* viene generalizzata a una misura su proiezioni in un reticolo non distributivo. La logica probabilistica classica è un caso limite: quando $\frac{\mathbb{M}}{\sim}$ è commutativa, $\text{Proj}(\frac{\mathbb{M}}{\sim})$ è isomorfo a una σ -algebra di sottoinsiemi di uno spazio, e Gleason collassa al teorema di rappresentazione classico. In altre parole, la probabilità quantistica è la generalizzazione più diretta della misura logica a contesti non commutativi.

La categoria degli spazi di Hilbert con canali quantistici completamente positivi e unital è, a sua volta, un'analoga “categoria di Markov quantistica”: i morfismi sono trasformazioni CPTP (completely positive trace-preserving). In questa categoria, la monade di Girard cede il posto alla struttura “state-effect”: ogni oggetto ha un compatto convesso di stati e un effetto-algebra di effetti; le probabilità sono valutazioni $q(E)$ di effetti negli stati. La logica probabilistica qui diventa logica degli effetti.

Probabilità classica, probabilità quantistica, e logica probabilistica si allineano in una scala unica: classica (reticoli booleani), intuitionista (Heyting, topos), quantistica (ortomodulari non distributivi). In tutti i casi, la probabilità è una funzione che rispetta la struttura d'ordine, ma la struttura stessa cambia.

Punti fissi probabilistici e semantica denotazionale avanzata

Nel Volume II, la logica temporale probabilistica PCTL e il probabilistic μ -calculus sono apparsi come linguaggi per descrivere proprietà di sistemi dinamici probabilistici, con semantica definita tramite misure su cammini e sistemi lineari o punti fissi.

Nel Volume III, si può guardare a questo con più profondità: i punti fissi probabilistici sono una manifestazione di una struttura più generale, in cui la probabilità è combinata con l'ordine e la continuità. La semantica denotazionale classica dei linguaggi imperativi si appoggia su domini CPO (complete partial orders), in cui le funzioni continue hanno punti fissi minimi (teorema di Kleene–Tarski). La controparte probabilistica richiede oggetti che siano contemporaneamente spazi misurabili e domini ordinati.

Un approccio è considerare domini di probabilità: insiemi D dotati sia di una topologia (o σ -algebra) sia di un ordine parziale, tali che $G(D)$ —lo spazio delle misure su D —sia ancora un dominio, con un ordine stocastico (per esempio, l'ordine di convex stochastic dominance). Un kernel di Markov $f : D \rightarrow G(D)$ genera un operatore T su $G(D)$, e le proprietà di lungo periodo sono punti fissi di T . Il teorema di Banach sui contrazioni, il teorema di Tarski sui punti fissi, e la teoria dei semigrupp di Markov si intrecciano: la probabilità diventa una dinamica su uno spazio di misure, e la logica probabilistica descrive questi punti fissi.

La formula $\mu X. \phi(X)$ nel probabilistic μ -calculus corrisponde a un punto fisso minimo di un operatore monotono F su $[0,1]^S$, dove S è l'insieme degli stati. Il fatto che tale punto fisso esista è garantito dall'ordine completo di $[0,1]^S$ e dalla monotonicità di F ; la sua unicità dipende spesso da contrazionalità in una metrica tipo Kantorovich. Questa è la stessa struttura che appare nella teoria dei markov decision processes (MDP), dove la value function V soddisfa un'equazione di Bellman:

$$V(s) = \max_a [r(s,a) + \gamma \sum_t P(s,a,t) V(t)].$$

La soluzione è il punto fisso di un operatore di Bellman che è una contrazione in norma sup; la convergenza iterativa è garantita dal teorema di Banach.

Tutto questo può essere riformulato in termini coalgebrici: un sistema dinamico probabilistico è una coalgebra F -struttura

$$c : S \rightarrow F(S),$$

dove $F(X)=G(A \times X)$ per qualche insieme di azioni A . Le proprietà sono descritte da logiche coalgebriche con modali probabilistici. Un μ -calcolo probabilistico è allora una logica di punti fissi per coalgebre. La teoria astratta dei punti fissi su lattice complete e le tecniche di coinduzione probabilistica diventano lo strato metalogico: la logica probabilistica è una sintassi per punti fissi di operatori stocastici su domini.

Entropia, informazione e semantica logico-informazionale

Fino ad ora, la probabilità è stata trattata come misura di “quasi ovunque verità” o di frequenza. Ma la teoria dell'informazione suggerisce una lettura diversa: la probabilità è una misura di sorpresa, e l'entropia è una misura di incertezza residua. In un certo senso, l'informazione è un'altra metrica sulla verità: una formula che riduce molto l'entropia porta molta informazione, una formula che riduce poco l'entropia porta poca informazione.

Dato uno spazio di modelli (M, μ) , ogni formula φ definisce un sottoinsieme $[\varphi] \subseteq M$. La probabilità $\Pr(\varphi) = \mu([\varphi])$ è il grado di verità statistica. Ma si può anche definire il guadagno di informazione nell'apprendere φ :

$$I(\varphi) = H(\mu) - H(\mu|\varphi),$$

dove $H(\mu)$ è l'entropia di μ , e $\mu|\varphi$ è la misura condizionata su $[\varphi]$. Questa quantità misura quanto restringere lo spazio dei modelli a quelli che soddisfano φ riduce l'incertezza epistemica. La logica probabilistica può allora essere vista come una logica dell'informazione: oltre a dire “ φ è molto probabile”, possiamo dire “ φ è molto informativa”.

Questo porta a nozioni logiche come: φ è ridondante rispetto a una teoria T se $I(\varphi | T) \approx 0$, cioè se l'aggiunta di φ a T non riduce l'entropia della distribuzione sui modelli della teoria. La compressione minima di descrizioni (Minimum Description Length, MDL) è un criterio che seleziona teorie T che minimizzano la lunghezza di codifica del modello + dei dati. Dal punto di vista logico, MDL è una logica preferenziale: tra le teorie che spiegano i dati, si preferiscono quelle con minore capacità descrittiva. La probabilità entra in MDL tramite il codice ottimale di Shannon–Fano: la lunghezza del codice per un modello M è approssimativamente $-\log \mu(M)$. Una teoria che massimizza $\mu(M)$ per i dati osservati minimizza la lunghezza di codifica.

Questo collega la logica probabilistica con una logica informazionale: le formule non sono solo vere o false, né solo probabili o improbabili, ma più o meno compressive. L'efficienza della spiegazione diventa un criterio logico.

Metriche probabilistiche, distanza tra modelli e coalgore metriche

Nel Volume I, la probabilità era una misura; nel Volume II, la distanza tra processi probabilistici era implicitamente la distanza statistica (total variation). Nel Volume III, si può astrarre questa idea: una logica probabilistica induce una metrica tra modelli, e questa metrica può essere caratterizzata in termini di bisimulazione quantitativa.

La distanza di Kantorovich (o Wasserstein) tra due misure μ, ν su uno spazio (X, d) è definita come:

$$W(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int d(x, y) d\pi(x, y),$$

dove $\Pi(\mu, \nu)$ è l'insieme dei couplings con margini μ e ν . Nel caso discreto, è il costo minimo per trasportare μ in ν . Questa distanza induce un sollevamento $\hat{F}$ di un funtore F : ad esempio, se $F(X) = G(X)$ (misure su X), si ottiene un funtore su $\mathbf{Met}$ (categorie di spazi metrici) che manda (X, d) in $(G(X), W)$. Un sistema probabilistico $S \rightarrow G(S)$ diventa una coalgebra in $\mathbf{Met}$, e si può definire una distanza bisimulativa come il più grande pseudometrico d su S tale che la mappa di transizione sia una funzione non espansiva rispetto a W .

Questa distanza quantifica quanto due stati siano comportamentalmente “diversi” dal punto di vista probabilistico. La logica probabilistica modale induce un'altra distanza:

$$d_L(s, t) = \sup_{\varphi} | \llbracket \varphi \rrbracket(s) - \llbracket \varphi \rrbracket(t) |,$$

dove $\llbracket \varphi \rrbracket(s)$ è il valore (probabilità, grado) della formula φ nello stato s . Sotto ipotesi di completezza logica, d_L coincide con la distanza bisimulativa. Questa è una forma quantitativa del principio di

Hennessy–Milner: due stati sono logicamente indistinguibili se e solo se la loro distanza logica è zero; più è grande la distanza, più formule li distinguono fortemente.

Questo rende esplicito qualcosa che nei volumi precedenti era implicito: una logica probabilistica non è solo un insieme di formule vere, ma un modo per misurare differenze di comportamento. La costruzione coalgebrica–metrica generalizza a molti funtori F : non solo Markov chains, ma sistemi probabilistici con input, output, pesi, ecc. La logica probabilistica corrispondente è una logica coalgebrica quantitativa.

Probabilità nei topos e logica interna

L'idea di fondo è ribaltare la situazione del Volume I: lì la logica viveva “sopra” la probabilità (formule $\rightarrow$ insiemi $\rightarrow$ misura). Qui invece la probabilità viene integrata nella **logica interna** di un topos. La domanda diventa: che cosa significa “misura di probabilità” se l'universo di discorso non è più **Set** ma un topos qualsiasi, con una logica intuizionista intrinseca?

1. Topos, logica interna e verità intuizionistica

Un topos elementare $\mathbb{T}$ è, molto sinteticamente, una categoria che si comporta come **Set** ma con logica intuizionista: è cartesianamente chiusa, ha oggetto terminale, oggetto dei sottoinsiemi (subobject classifier) Ω , pullback stabili, ecc.

Il ruolo cruciale lo gioca Ω : per ogni monomorfismo (inclusione) $m: U \hookrightarrow X$, esiste una mappa caratteristica $\chi_m: X \rightarrow \Omega$ tale che:

- “ x appartiene a U ” è espresso internamente da $\chi_m(x) = \top$ in Ω ;
- Ω , con le sue operazioni interne, forma un'algebra di Heyting, non necessariamente booleana.

Quindi:

- in **Set**, $\Omega = \{0,1\}$, e la logica interna è classica;
- in un topos generale, Ω può avere verità “parziali”, e la logica interna è intuizionista.

La semantica alla Kripke–Joyal ci dice cosa significa, internamente, che una formula ϕ è vera in un oggetto X : non è una proprietà assoluta, ma dipende da “stati di informazione”, cioè morfismi $U \rightarrow X$.

Ora, all'inizio del paper, una misura di probabilità era una funzione $\mu: \Sigma \rightarrow [0,1]$ su una σ -algebra di sottoinsiemi. In un topos, i “sottoinsiemi” diventano **suboggetti**, classificati da mappe a Ω . Dunque, se voglio una “misura interna”, devo assegnare a ogni suboggetto $U \hookrightarrow X$ un valore in un oggetto che gioca il ruolo di $[0,1]$, ma rispettando la logica intuizionista.

2. Oggetti di probabilità interni: da $[0,1]$ a quantali interni

Per imitare $[0,1]$ in un topos, si può costruire un oggetto $\mathbb{P}$ con struttura di intervallo:

- oggetto interno con ordinamento $\leq$,
- operazioni interne $+$, $-$, $\wedge$, $\vee$,
- elementi distinti $0,1$.

Ma in generale, per trattare additività e combinazione di eventi, è più naturale pensare a $\mathbb{I}$ come a un **quantale commutativo interno**: un monoide commutativo $(\mathbb{I}, \oplus, 0)$ con ordine $\leq$ tale che $\oplus$ distribuisce sui $\sup$ ($\vee$). Questo permette di scrivere, internamente:

- additività finita: $\mu(U \amalg V) = \mu(U) \oplus \mu(V)$ quando U e V sono disgiunti;
- monotonia: $U \leq V \Rightarrow \mu(U) \leq \mu(V)$.

La σ -additività richiede di parlare di somme numerabili: nella logica interna del topos, si può esprimere come proprietà per famiglie indicizzate da $\mathbb{N}$ (l'oggetto interno dei naturali).

In **Set**, un quantale commutativo con certi assiomi è isomorfo a $[0, 1]$ con $+$ troncata o con la somma usuale sotto vincoli. In un topos, $\mathbb{I}$ gioca il ruolo di “valori di probabilità interni”.

3. Misure interne su oggetti: μ come trasformazione naturale

Dato un oggetto X , la nozione naturale di “ σ -algebra” interna è l'insieme dei suboggetti di X stabili sotto certe operazioni interne (un sotto-Heyting-sottooggetto dell'algebra $\text{Sub}(X)$).

Una **misura interna su X** è allora:

- oppure come trasformazione naturale $\mu_X : \text{Sub}(X) \rightarrow \mathbb{I}$, monotona e additiva su unioni disgiunte;
- oppure, dualmente, come mappa di tipo $\mu : X \rightarrow G(X)$ dove G è una monade di probabilità interna (estensione della Giry monad a $\mathbb{I}\mathbf{hLab}$).

In categoria di Kleisli di Giry in **Set**, i morfismi $X \rightarrow G(Y)$ sono kernel di Markov. In un topos $\mathbb{I}$, una **monade di probabilità interna** è un endofuntore

$$G : \mathbb{I} \rightarrow \mathbb{I}$$

dotato di:

- unità $\eta_X : X \rightarrow G(X)$,
- moltiplicazione $\mu_X : G(G(X)) \rightarrow G(X)$,

che soddisfano le equazioni monadiche, e la cui interpretazione interna è: per ogni X , $G(X)$ è “lo spazio delle misure interne su X ” $\mathbf{nLab}$.

Un punto chiave: in molti casi di interesse (topos di sheaves su uno spazio misurabile, topos di presheaves su una categoria di “stati”), G può essere costruita “punto per punto”: $G(X)(U) = G(X(U))$ per ogni oggetto di base U della categoria sottostante. Quindi la probabilità interna è “locale”: per ogni condizione U , si ha una misura su $X(U)$, e l'adesione alle leggi di gluing garantisce coerenza globale.

In questa prospettiva, i quantificatori probabilistici del Volume I, del tipo $(P \ x \geq r) \ \varphi(x)$, diventano **formule della logica interna** che affermano:

- “il suboggetto di X definito da φ ha misura $\geq r$ ”,

dove r non è un numero reale esterno, ma un elemento di $\mathbb{I}$ interno.

4. Logica interna, quantificatori probabilistici e semantica di tipo Gaifman generalizzata

Nel Volume I, la semantica alla Gaifman diceva: una misura μ su uno spazio di modelli Q e un predicato $\varphi(x)$ definiscono il set $\{M \in Q \mid M \models \varphi(a)\}$ e la sua misura. In un topos, questa idea si internalizza: l'oggetto dei modelli M (ad esempio un oggetto in $\mathbb{M}$ che classifica modelli di una teoria) ha suboggetti definibili da formule interne, e una misura interna $\mu_M : \text{Sub}(M) \rightarrow \mathbb{P}$ assegna probabilità interne alle classi definibili.

Una formula con quantificatore probabilistico interna avrà forma, concettualmente:

$$(P x \geq r) \varphi(x)$$

e verrà interpretata come:

$$\mu_X(\{x \mid \varphi(x)\}) \geq r,$$

dove:

- $\{x \mid \varphi(x)\}$ è un suboggetto interno di X definito nella logica interna del topos;
- μ_X è una misura interna su X ;
- r è un elemento di $\mathbb{P}$ (possibilmente non standard rispetto a $\mathbb{R}$ classici).

Questa lettura cambia radicalmente lo “status ontologico” della probabilità: non è più un oggetto esterno (numeri reali) a cui si mappa, ma una componente interna del medesimo universo logico. La distinzione tra “logica” e “probabilità” si attenua: la probabilità è un funtore interno che arricchisce la logica.

5. Topos di sheaves come modelli di probabilità locale

Un caso estremamente istruttivo è il topos di sheaves su uno spazio topologico T (o su un sito), $\text{Sh}(T)$.

Ogni oggetto F in $\text{Sh}(T)$ è un fascio: ad ogni aperto U di T associa un insieme $F(U)$, con restrizioni coerenti.

Si può pensare a T come a uno spazio di contesti, “stati di informazione” o “stadi di osservazione”. Un fascio F rappresenta un sistema la cui configurazione dipende dal contesto: $F(U)$ sono le configurazioni viste restringendosi a U .

Una misura interna su F dovrebbe assegnare, per ogni U , una misura μ_U su $F(U)$, in modo compatibile con le restrizioni: se $V \subseteq U$, allora il pushforward della misura μ_U sotto la restrizione $F(U) \rightarrow F(V)$ è μ_V . In formula, per ogni suboggetto definito da una formula φ su $F(V)$:

$$\mu_V([\varphi]_V) = \mu_U([\varphi]_U \cap \text{res}^{-1}([\varphi]_V)).$$

Questa è una versione sheaf-theoretic della **condizionalizzazione**: le misure sui contesti più grandi determinano le misure sui contesti più piccoli, e viceversa si può vedere μ come un fascio di misure. Strutture categoriali recenti come le Markov categories e i probability monads vengono reinterpretate in contesti di questo tipo [arXiv.org+1](https://arxiv.org/abs/1908.07251).

La logica interna di $\text{Sh}(\mathcal{T})$ è intuizionista; quindi, la probabilità interna in $\text{Sh}(\mathcal{T})$ è automaticamente compatibile con una logica non classica. Questo permette di modellare situazioni in cui:

- la verità di φ non è ancora decidibile globalmente (φ non è né vera né falsa, ma localmente vera su alcuni U e falsa su altri);
- la probabilità di φ può essere definita localmente, e la composizione globale è un'integrazione “su $\mathcal{T}$ ”.

Probabilità quantistica e logica quantistica

La probabilità quantistica è, a tutti gli effetti, la più profonda generalizzazione della probabilità classica mai formulata. Non è una variante “strana” della probabilità, ma una teoria coerente, completa, e sorprendentemente naturale, che emerge inevitabilmente quando si abbandona la commutatività degli eventi. In un certo senso, la probabilità quantistica è ciò che si ottiene quando si prende sul serio l'idea che gli eventi non siano sempre compatibili, e che l'ordine delle osservazioni possa influire sui risultati.

Il punto di partenza è semplice: nella probabilità classica, gli eventi sono sottoinsiemi di uno spazio Ω , e la probabilità è una misura $\mu: \Sigma \rightarrow [0,1]$. Nella probabilità quantistica, gli eventi non sono più sottoinsiemi, ma **proiezioni** in uno spazio di Hilbert. La probabilità non è più una misura su insiemi, ma uno **stato** su un'algebra di operatori. La logica degli eventi non è più booleana, ma **ortomodulare**.

Questa trasformazione è radicale: la struttura logica cambia, la struttura geometrica cambia, la struttura probabilistica cambia. E tuttavia, tutto rimane perfettamente coerente.

1. Eventi come proiezioni: la struttura ortomodulare

Sia $\mathcal{H}$ uno spazio di Hilbert complesso separabile. Una **proiezione** è un operatore $P: \mathcal{H} \rightarrow \mathcal{H}$ tale che:

$$P = P^* = P^2.$$

L'insieme $\text{Proj}(\mathcal{H})$ di tutte le proiezioni è un reticolo parzialmente ordinato da:

$$P \leq Q \Leftrightarrow PQ = P.$$

L'ortocomplemento è:

$$P^\perp = I - P.$$

La congiunzione e la disgiunzione non sono più semplici intersezioni e unioni, ma:

$P \wedge Q$ = proiezione sull'intersezione dei sottospazi, $P \vee Q$ = proiezione sulla chiusura della somma dei sottospazi.

Il reticolo $\text{Proj}(\mathcal{H})$ è:

- ortomodulare,
- non distributivo,
- atomico,

- completo.

La non distributività è il cuore della logica quantistica:

$$P \wedge (Q \vee R) \neq (P \wedge Q) \vee (P \wedge R)$$

in generale. Questo è il riflesso logico della non commutatività degli osservabili.

2. Stati quantistici: probabilità non commutative

In un'algebra di von Neumann $\frac{\mathcal{M}}{\mathcal{K}} \subseteq B(\mathcal{H})$, uno **stato** è una funzione:

$$\varrho : \frac{\mathcal{M}}{\mathcal{K}} \rightarrow \mathbb{C}$$

che soddisfa:

1. linearità: $\varrho(aA + bB) = a\varrho(A) + b\varrho(B)$,
2. positività: $\varrho(A^* A) \geq 0$,
3. normalizzazione: $\varrho(I) = 1$,
4. continuità σ -debole (normalità).

Se ρ è un operatore densità (positivo, trace-class, $\text{tr}(\rho)=1$), allora:

$$\varrho(A) = \text{tr}(\rho A).$$

La probabilità di un evento $P \in \text{Proj}(\frac{\mathcal{M}}{\mathcal{K}})$ è:

$$\text{Pr}(P) = \varrho(P) = \text{tr}(\rho P).$$

Questo è il cuore della probabilità quantistica: la probabilità è una valutazione su proiezioni, non su insiemi.

3. Il teorema di Gleason: la struttura inevitabile della probabilità quantistica

Il teorema di Gleason (1957) è uno dei risultati più profondi della matematica del XX secolo. Afferma che:

In dimensione ≥ 3 , ogni misura finitamente additiva su $\text{Proj}(\mathcal{H})$ è della forma $\text{tr}(\rho P)$ per un unico operatore densità ρ .

Formalmente:

Se $\mu: \text{Proj}(\mathcal{H}) \rightarrow [0,1]$ soddisfa:

- $\mu(I)=1$,
- $\mu(P_1 + \dots + P_n) = \mu(P_1) + \dots + \mu(P_n)$ per proiezioni ortogonali,

allora esiste un unico ρ tale che $\mu(P)=\text{tr}(\rho P)$.

Questo risultato è straordinario per due motivi:

1. **La probabilità quantistica non è una scelta: è imposta dalla geometria di $\mathcal{H}$.** Non esistono altre misure coerenti.
2. **La probabilità quantistica è più rigida della probabilità classica.** In un reticolo booleano, esistono infinite misure; in $\text{Proj}(\mathcal{H})$, tutte derivano da stati densità.

Il teorema di Gleason è la ragione per cui la meccanica quantistica è una teoria probabilistica inevitabile: la struttura logica degli eventi impone la forma della probabilità.

4. Logica quantistica: ortomodularità e fallimento della distributività

La logica quantistica nasce dal tentativo di interpretare $\text{Proj}(\mathcal{H})$ come algebra di proposizioni. Le operazioni logiche sono:

- negazione: $P \perp$,
- congiunzione: $P \wedge Q$,
- disgiunzione: $P \vee Q$.

Il reticolo è ortomodulare:

$$P \leq Q \Rightarrow Q = P \vee (Q \wedge P \perp).$$

Ma non è distributivo. Questo ha conseguenze profonde:

- non esiste una logica classica che descriva gli eventi quantistici;
- la legge del terzo escluso non vale in generale;
- la logica quantistica è più vicina a una logica geometrica che a una logica booleana.

La logica quantistica è stata formalizzata da Birkhoff e von Neumann nel 1936. È una logica proposizionale non classica, con regole di inferenza modificate per rispettare l'ortomodularità.

5. Stati, effetti e logica degli effetti

La formulazione moderna della probabilità quantistica non usa solo proiezioni, ma **effetti**: operatori E tali che $0 \leq E \leq I$. Gli effetti rappresentano eventi “sfumati”, non necessariamente proiettivi.

La probabilità di un effetto E in uno stato ρ è:

$$\text{Pr}(E) = \text{tr}(\rho E).$$

Gli effetti formano un **effetto-algebra**, una struttura parzialmente definita con operazione $\oplus$ (somma parziale) e ortocomplemento $E \perp = I - E$.

La logica degli effetti è una logica probabilistica generalizzata:

- gli eventi non sono più veri/falsi,
- ma hanno un valore in $[0,1]$ determinato dallo stato.

Questo è il ponte tra probabilità classica, fuzzy logic e probabilità quantistica.

6. Probabilità quantistica come caso particolare di probabilità non commutativa

La probabilità quantistica è un caso particolare della teoria delle **algebre di von Neumann** e degli **stati normali**. In generale, una probabilità non commutativa è uno stato su un'algebra $\mathfrak{A}$

$$\varrho: \mathfrak{A} \rightarrow \mathbb{C}.$$

La probabilità classica è il caso in cui $\mathfrak{A}$ è commutativa: $\mathfrak{A} \cong L^\infty(\Omega, \Sigma, \mu)$. La probabilità quantistica è il caso in cui $\mathfrak{A}$ è non commutativa: $\mathfrak{A} \cong B(\mathcal{H})$.

Questa unificazione è uno dei trionfi della teoria operatoriale: la probabilità è una teoria degli stati su algebre, e la commutatività o meno dell'algebra determina la logica degli eventi.

7. Misure condizionate e aggiornamento bayesiano quantistico

Nella probabilità classica, l'aggiornamento bayesiano è:

$$\mu(A|B) = \mu(A \wedge B) / \mu(B).$$

Nella probabilità quantistica, l'aggiornamento dopo una misura proiettiva P è:

$$\rho \mapsto \rho' = P\rho P / \text{tr}(\rho P).$$

Questo è il **postulato di Lüders**. È una forma di condizionamento non commutativo. Se P e Q non commutano, l'ordine delle misure conta:

$$P(Q(\rho)) \neq Q(P(\rho)).$$

Questo è il riflesso probabilistico della non commutatività logica.

8. Logiche modali quantistiche e bisimulazione quantistica

Nel Volume II, la bisimulazione probabilistica era definita su kernel di Markov. Nel contesto quantistico, i processi sono canali CPTP (completely positive trace-preserving):

$$\Phi: \rho \mapsto \sum_i K_i \rho K_i^*.$$

Una **bisimulazione quantistica** è una relazione tra stati densità che preserva:

- probabilità di effetti,
- evoluzione tramite canali,
- equivalenza operativa.

La logica modale quantistica ha formule del tipo:

$\langle E \rangle \geq p$

che significa: “la probabilità dell’effetto E è $\geq p$ ”.

La semantica è:

$$\rho \models \langle E \rangle \geq p \Leftrightarrow \text{tr}(\rho E) \geq p.$$

Questa logica è completa rispetto alla bisimulazione quantistica.

9. Il ruolo del teorema di Kochen–Specker

Il teorema di Kochen–Specker (1967) afferma che:

Non esiste una valutazione classica (0/1) su $\text{Proj}(\mathcal{H})$ che preservi le relazioni logiche, se $\dim \mathcal{H} \geq 3$.

Questo significa che:

- non esiste una probabilità classica nascosta che riproduca la probabilità quantistica;
- la logica quantistica non è un reticolo booleano mascherato;
- la non commutatività è intrinseca.

Il teorema di Kochen–Specker è il complemento logico del teorema di Gleason: insieme, mostrano che la probabilità quantistica è inevitabile.

10. Sintesi concettuale

La probabilità quantistica è:

- una probabilità su proiezioni, non su insiemi;
- una misura non commutativa, non una misura classica;
- una logica ortomodulare, non una logica booleana;
- una teoria di stati su algebre, non una teoria di misure su spazi;
- una teoria di aggiornamento non commutativo, non un semplice Bayes;
- una teoria di processi CPTP, non di kernel di Markov.

È, in un certo senso, la generalizzazione più naturale e più profonda della probabilità classica.

La semantica dei sistemi probabilistici, quando la si osserva con sufficiente profondità, rivela una struttura che non è più quella di una semplice misura su insiemi di cammini, ma quella di un punto fisso di un operatore stocastico definito su uno spazio di funzioni o su uno spazio di misure. Ogni proprietà di lungo periodo, ogni comportamento asintotico, ogni probabilità di raggiungimento, ogni valore atteso, ogni formula del μ -calcolo probabilistico è, in ultima analisi, la soluzione di un’equazione di punto fisso. Questa osservazione, apparentemente tecnica, è in realtà la chiave per comprendere l’unità profonda tra logica, probabilità e dinamica. Nel Volume II, la logica temporale probabilistica PCTL e il probabilistic μ -calculus apparivano come linguaggi per descrivere proprietà di sistemi dinamici probabilistici; qui, nel Volume III, si mostra che tali linguaggi non sono semplici strumenti descrittivi, ma la sintassi naturale di una teoria dei punti fissi probabilistici su domini ordinati e strutture coalgebriche.

Consideriamo un sistema probabilistico discreto, una catena di Markov con insieme di stati S e matrice di transizione $P(s,t)$. Una proprietà come “la probabilità di raggiungere un insieme T di stati” è definita come la soluzione della seguente equazione: $x(s) = 1$ se $s \in T$, $x(s) = \sum_t P(s,t) x(t)$ se $s \notin T$. Questa è un’equazione di Bellman, e la sua soluzione è il punto fisso minimo dell’operatore $F(x)(s) = 1_T(s) + (1 - 1_T(s)) \sum_t P(s,t) x(t)$. L’esistenza del punto fisso è garantita dal fatto che $[0,1]^S$ è un dominio completo, ordinato punto per punto, e F è monotono. Il teorema di Tarski assicura che F ha un punto fisso minimo, ottenibile come limite della catena crescente $F^0(\perp)$, $F^1(\perp)$, $F^2(\perp)$, ..., dove $\perp$ è la funzione costante zero. Questo è esattamente ciò che accade quando si calcola iterativamente la probabilità di raggiungimento: si parte da zero e si propaga la probabilità lungo la dinamica.

Ma questa struttura non è un accidente: è la manifestazione di un principio generale. Ogni proprietà probabilistica definibile da un μ -calcolo quantitativo è un punto fisso di un operatore monotono su un dominio completo. Una formula $\mu X. \varphi(X)$ è interpretata come il punto fisso minimo della funzione F definita da φ ; una formula $\nu X. \varphi(X)$ è il punto fisso massimo. La semantica del μ -calcolo probabilistico è quindi una semantica di punti fissi, e la logica stessa è una sintassi per descrivere operatori monotoni su domini probabilistici.

Il passo successivo è riconoscere che gli stati di un sistema probabilistico non sono semplicemente elementi di un insieme S , ma elementi di una coalgebra. Una coalgebra per un funtore F è una mappa $c: S \rightarrow F(S)$. Nel caso probabilistico, il funtore è $F(X) = G(A \times X)$, dove G è la monade di Giry o la sua versione discreta, e A è un insieme di azioni. Una coalgebra probabilistica è quindi una funzione che, dato uno stato, restituisce una distribuzione sulle coppie (azione, stato successivo). La dinamica probabilistica è una coalgebra, e le proprietà sono invarianti coalgebrici. La logica modale probabilistica è una logica coalgebrica, e il μ -calcolo probabilistico è una logica di punti fissi per coalgebre. La bisimulazione probabilistica è la relazione di equivalenza naturale per coalgebre probabilistiche: due stati sono bisimili se hanno la stessa struttura di transizione, e la logica è completa rispetto alla bisimulazione.

Ma la coalgebra non è solo una struttura sintattica: è anche una struttura metrica. Ogni sistema probabilistico induce una distanza tra stati, definita come la distanza logica: $d(s,t) = \sup_{\varphi} |\llbracket \varphi \rrbracket(s) - \llbracket \varphi \rrbracket(t)|$, dove φ varia su tutte le formule della logica probabilistica quantitativa. Questa distanza è un pseudometrico, e misura quanto due stati sono distinguibili da osservatori probabilistici. La distanza logica coincide con la distanza bisimulativa quantitativa, definita come il più grande pseudometrico d tale che la mappa di transizione sia non espansiva rispetto alla distanza di Kantorovich sollevata. La distanza di Kantorovich tra due misure μ e ν su uno spazio metrico (X,d) è definita come $W(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int d(x,y) d\pi(x,y)$, dove $\Pi(\mu,\nu)$ è l’insieme dei couplings. Questa distanza solleva la metrica da X a $G(X)$, e permette di definire una distanza tra stati di una coalgebra probabilistica. Il risultato fondamentale è che la distanza logica e la distanza bisimulativa coincidono: la logica è espressivamente completa rispetto alla metrica.

Questa struttura metrica permette di interpretare i punti fissi probabilistici come limiti di contrazioni. Molti operatori probabilistici, come l’operatore di Bellman nei processi decisionali di Markov, sono contrazioni in norma sup o in distanza di Kantorovich. Il teorema di Banach assicura l’unicità del punto fisso e la convergenza iterativa. Questo è il cuore della teoria del controllo stocastico, dell’apprendimento per rinforzo e della semantica quantitativa dei programmi probabilistici. La probabilità non è più solo una misura, ma una dinamica su uno spazio metrico di funzioni o di misure, e le proprietà sono punti fissi di operatori contrattivi.

La teoria dei domini entra in gioco quando si considera la continuità degli operatori. Un dominio è un insieme parzialmente ordinato in cui ogni catena crescente ha un supremo. Gli operatori continui

preservano i supremi delle catene, e quindi hanno punti fissi ottenibili come limiti di iterazioni. Nel caso probabilistico, gli operatori sono spesso continui rispetto all'ordine punto per punto, e quindi i punti fissi esistono e sono calcolabili come limiti di approssimazioni. La semantica denotazionale dei programmi probabilistici è costruita su domini di funzioni da S a $[0,1]$, ordinati punto per punto, e gli operatori semantici sono continui. Il μ -calcolo probabilistico è quindi una logica di domini, e la sua semantica è un caso particolare della teoria dei punti fissi su CPO.

La coalgebra fornisce la struttura sintattica, la teoria dei domini fornisce la struttura ordinata, la metrica di Kantorovich fornisce la struttura quantitativa, e il μ -calcolo probabilistico fornisce la sintassi dei punti fissi. Tutto converge in un'unica visione: un sistema probabilistico è una coalgebra in una categoria metrica, le sue proprietà sono punti fissi di operatori monotoni o contrattivi, e la logica probabilistica è la sintassi naturale per descrivere tali punti fissi. La probabilità, vista da qui, non è più una misura statica, ma una dinamica di approssimazione, un processo di convergenza, una struttura di punto fisso.

In questa prospettiva, il Volume III non è un'estensione del Volume II, ma la sua rivelazione strutturale. Ciò che nel Volume II appariva come un insieme di tecniche — equazioni di Bellman, probabilità di raggiungimento, μ -calcolo, bisimulazione — qui si mostra come manifestazione di un'unica architettura matematica: la probabilità come punto fisso di un operatore stocastico su un dominio metrico coalgebrico. La logica probabilistica è la grammatica di questa architettura; la teoria dei domini è la sua topologia; la metrica di Kantorovich è la sua geometria; la coalgebra è la sua sintassi; il μ -calcolo è la sua teoria dei punti fissi.

La probabilità, nella sua forma classica, misura la frequenza o la credenza; ma nella teoria dell'informazione di Shannon, la probabilità diventa una misura di sorpresa, e l'entropia diventa una misura di incertezza residua. Questa reinterpretazione è molto più che un cambio di prospettiva: è un cambio di ontologia. La probabilità non è più un numero che quantifica quanto un evento è plausibile, ma un numero che quantifica quanto un evento è informativo. L'entropia $H(\mu) = -\sum \mu(x) \log \mu(x)$ non è una proprietà del mondo, ma una proprietà della nostra conoscenza del mondo. In questa prospettiva, la logica probabilistica non è più una logica della verità quasi ovunque, ma una logica dell'informazione: una logica che misura quanto una formula riduce l'incertezza epistemica.

Consideriamo uno spazio di modelli M , con una misura μ che rappresenta la nostra credenza iniziale. Ogni formula ϕ definisce un sottoinsieme $[\phi] \subseteq M$, e la probabilità $\Pr(\phi) = \mu([\phi])$ è la credenza che ϕ sia vera. Ma la quantità più interessante non è $\Pr(\phi)$, bensì il guadagno di informazione nell'apprendere ϕ : $I(\phi) = H(\mu) - H(\mu|\phi)$, dove $\mu|\phi$ è la misura condizionata su $[\phi]$. Questa quantità misura quanto la formula ϕ restringe lo spazio dei modelli compatibili con la nostra conoscenza. Una formula con alta probabilità può essere poco informativa; una formula con bassa probabilità può essere molto informativa. La logica classica non distingue tra queste due dimensioni; la logica probabilistica informazionale sì.

Il guadagno di informazione è una misura di compressione: se una teoria T implica ϕ , allora la descrizione di T deve contenere almeno $I(\phi)$ bit di informazione. Questo collega la logica probabilistica alla teoria della complessità descrittiva. Una teoria T è più "semplice" di una teoria T' se la sua descrizione è più corta, e la lunghezza ottimale di una descrizione è proporzionale all'entropia della distribuzione dei modelli che la teoria lascia aperti. La teoria del Minimum Description Length (MDL) formalizza questa idea: la migliore teoria è quella che minimizza $L(T) + L(D|T)$, dove $L(T)$ è la lunghezza della descrizione della teoria e $L(D|T)$ è la lunghezza della descrizione dei dati dato T . Ma $L(D|T)$ è, in media, l'entropia condizionata $H(\mu|T)$. Quindi MDL è una logica probabilistica informazionale: preferisce le teorie che comprimono meglio i dati.

In questa prospettiva, la probabilità non è più un fine, ma un mezzo: serve a definire una metrica informazionale sui modelli. Due modelli M e N sono vicini se la loro distanza informazionale è piccola, cioè se la loro distribuzione di probabilità sugli eventi osservabili è simile. La distanza naturale è la divergenza di Kullback–Leibler: $D(\mu||\nu) = \sum \mu(x) \log(\mu(x)/\nu(x))$. Questa quantità non è simmetrica, ma misura quanto la distribuzione μ è inefficiente se codificata con un codice ottimale per ν . In logica probabilistica, $D(\mu||\nu)$ misura quanto un osservatore che crede in μ sarebbe sorpreso se il mondo fosse governato da ν . La divergenza è quindi una misura di dissimilarità epistemica.

La divergenza induce una nozione di implicazione informazionale: una teoria T implica una teoria T' se la divergenza tra le loro distribuzioni sui modelli è zero. Questo è un criterio più fine dell'implicazione logica classica: due teorie possono essere logicamente equivalenti ma informazionalmente diverse, se inducono distribuzioni diverse sui modelli. La logica informazionale distingue tra teorie che hanno lo stesso contenuto proposizionale ma diversa struttura probabilistica. Questo è essenziale in contesti come l'apprendimento automatico, dove due modelli possono essere logicamente equivalenti ma avere diversa capacità di generalizzazione.

La semantica informazionale può definire una nozione di indipendenza logica: due formule ϕ e ψ sono informazionalmente indipendenti se $I(\phi \wedge \psi) = I(\phi) + I(\psi)$. Questo è l'analogo informazionale dell'indipendenza probabilistica, ma applicato a formule logiche. La logica classica non ha una nozione naturale di indipendenza; la logica probabilistica informazionale sì. Questo permette di definire una teoria della modularità logica: una teoria è modulare se le sue parti sono informazionalmente indipendenti. La modularità è una proprietà cruciale per la comprensibilità e la manutenibilità delle teorie.

La semantica informazionale si intreccia con la semantica dei punti fissi del capitolo precedente. Un punto fisso probabilistico può essere interpretato come uno stato di equilibrio informazionale: una distribuzione μ tale che l'applicazione dell'operatore stocastico non cambia l'entropia. Questo è il caso delle distribuzioni stazionarie delle catene di Markov, che massimizzano l'entropia soggetta ai vincoli imposti dalla dinamica. La logica probabilistica può esprimere proprietà di massima entropia: una formula ϕ è una legge di massima entropia se la distribuzione che massimizza $H(\mu)$ soggetta a ϕ è la distribuzione “più naturale” compatibile con ϕ . Questo collega la logica probabilistica alla teoria delle inferenze di Jaynes, in cui la probabilità è la soluzione di un problema di ottimizzazione informazionale.

La semantica informazionale permette anche di reinterpretare la nozione di bisimulazione probabilistica. Due stati sono bisimili se nessuna formula probabilistica li distingue; ma si può raffinare questa nozione chiedendo che nessuna formula informazionale li distingua, cioè che la loro distanza informazionale sia zero. Questo porta alla bisimulazione informazionale, una relazione più fine della bisimulazione probabilistica, che tiene conto non solo delle probabilità, ma anche della struttura informazionale delle transizioni. In questa prospettiva, la logica probabilistica non è solo una logica della verità quantitativa, ma una logica della conoscenza quantitativa.

Infine, la semantica informazionale permette di reinterpretare la logica probabilistica come una teoria della compressione. Ogni formula ϕ è un vincolo che riduce lo spazio dei modelli, e quindi riduce l'entropia. La teoria migliore è quella che comprime meglio i dati, cioè quella che massimizza il guadagno di informazione per unità di complessità descrittiva. Questo è il principio di Solomonoff, secondo cui la probabilità di una teoria è proporzionale a $2^{-K(T)}$, dove $K(T)$ è la complessità di Kolmogorov della teoria. La logica probabilistica informazionale è quindi una logica bayesiana universale, in cui la probabilità è una misura della semplicità descrittiva.

In questa visione unificata, la probabilità, l'informazione, la logica e la dinamica non sono più discipline separate, ma aspetti di un'unica struttura concettuale. La probabilità misura la sorpresa, l'entropia misura l'incertezza, la logica misura la compatibilità, la dinamica misura la convergenza, e la semantica informazionale misura la compressione. Il Volume III mostra che queste nozioni non sono indipendenti, ma parti di un'unica architettura epistemica. La logica probabilistica non è solo una logica della verità, ma una logica della conoscenza, dell'informazione e della spiegazione.

La probabilità, quando viene osservata non come misura ma come trasformazione, rivela una struttura geometrica che non è immediatamente evidente nei Volumi I e II. Ogni sistema probabilistico induce una distanza tra stati, una distanza tra modelli, una distanza tra comportamenti. Questa distanza non è un artificio, ma la manifestazione quantitativa della logica probabilistica: due stati sono vicini se nessuna formula probabilistica li distingue significativamente, e sono lontani se esiste una formula che li separa con un margine elevato. La logica probabilistica, vista da questa prospettiva, non è più un linguaggio di verità, ma un linguaggio di distanze. Ogni formula è una funzione 1-Lipschitz, ogni operatore modale è una contrazione, ogni punto fisso è un equilibrio metrico.

Consideriamo un sistema probabilistico discreto, una catena di Markov con insieme di stati S . La logica probabilistica quantitativa assegna a ogni formula ϕ un valore reale $\llbracket \phi \rrbracket(s) \in [0,1]$ per ogni stato s . La distanza logica tra due stati s e t è definita come $d_L(s,t) = \sup_{\phi} |\llbracket \phi \rrbracket(s) - \llbracket \phi \rrbracket(t)|$. Questa distanza misura quanto un osservatore, dotato della logica probabilistica, può distinguere s da t . Se $d_L(s,t)=0$, nessuna formula distingue i due stati; se $d_L(s,t)=1$, esiste una formula che li separa completamente. Questa distanza è un pseudometrico, e la logica probabilistica è espressivamente completa rispetto a esso: ogni funzione 1-Lipschitz può essere approssimata da formule della logica.

La distanza logica non è un'invenzione ad hoc: coincide con la distanza bisimulativa quantitativa, definita come il più grande pseudometrico d tale che, per ogni coppia di stati s e t , la distanza tra le loro distribuzioni di transizione sia controllata dalla distanza tra gli stati successivi. Per formalizzare questa idea, si solleva la distanza d da S allo spazio delle misure $G(S)$ usando la distanza di Kantorovich: $W_d(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int d(x,y) d\pi(x,y)$, dove $\Pi(\mu, \nu)$ è l'insieme dei couplings con margini μ e ν . Questa distanza misura il costo minimo per trasformare μ in ν , dove il costo di spostare massa da x a y è $d(x,y)$. Il sollevamento di Kantorovich è la struttura naturale per confrontare distribuzioni in termini della distanza tra i loro punti di supporto.

Una coalgebra probabilistica è una mappa $c: S \rightarrow G(S)$. La distanza bisimulativa quantitativa è il più grande pseudometrico d tale che $d(s,t) \geq W_d(c(s), c(t))$. Questa equazione è un punto fisso: d è il punto fisso massimo dell'operatore $T(d)(s,t) = W_d(c(s), c(t))$. L'esistenza del punto fisso è garantita dal fatto che lo spazio dei pseudometrici su S è un dominio completo, e T è monotono. La distanza bisimulativa è quindi un punto fisso di un operatore contrattivo o monotono, e la logica probabilistica è la sintassi naturale per descrivere tale punto fisso.

La distanza di Kantorovich non è solo uno strumento tecnico: è la metrica naturale della probabilità. Se la probabilità è una misura di incertezza, la distanza di Kantorovich è una misura di dissimilarità tra incertezze. Se la probabilità è una trasformazione, la distanza di Kantorovich è una misura di quanto due trasformazioni differiscono. Se la probabilità è una dinamica, la distanza di Kantorovich è una misura di quanto due dinamiche divergono. La logica probabilistica, vista da qui, è una logica metrica: ogni formula è una funzione 1-Lipschitz, ogni operatore modale è una contrazione, ogni punto fisso è un equilibrio metrico.

Questa struttura metrica permette di reinterpretare la bisimulazione probabilistica come una nozione quantitativa: due stati non sono semplicemente equivalenti o non equivalenti, ma più o meno simili. La bisimulazione classica è un caso limite: due stati sono bisimili se la loro distanza è zero. La

distanza bisimulativa è una generalizzazione naturale della bisimulazione, che permette di misurare la differenza tra stati non perfettamente equivalenti. Questo è essenziale in contesti come l'apprendimento, il controllo stocastico, la verifica quantitativa, dove non si cerca un'equivalenza esatta, ma una approssimazione controllata.

La struttura coalgebrica metrica permette anche di definire una nozione di continuità comportamentale: un sistema probabilistico è robusto se piccole perturbazioni nelle probabilità di transizione inducono piccole variazioni nella distanza logica. Questo è il cuore della teoria della robustezza quantitativa: la logica probabilistica non è solo un linguaggio per descrivere proprietà, ma un linguaggio per misurare la sensibilità delle proprietà rispetto alle perturbazioni. La distanza logica è la metrica naturale per questa sensibilità.

La distanza di Kantorovich permette anche di definire una nozione di convergenza comportamentale: una sequenza di sistemi probabilistici converge se le loro coalgebre convergono in distanza. Questo è il fondamento della teoria dei limiti dei sistemi probabilistici, analoga alla teoria dei graphons per i grafi densi. La logica probabilistica è completa rispetto a questa convergenza: se una sequenza di sistemi converge in distanza bisimulativa, allora converge per ogni formula della logica. La logica probabilistica è quindi una logica di continuità: descrive proprietà che sono stabili sotto limiti.

La metrica probabilistica consente di reinterpretare la semantica dei punti fissi. Un punto fisso probabilistico non è solo una soluzione di un'equazione, ma un equilibrio metrico: un punto in cui la dinamica non cambia più la distanza. La convergenza iterativa verso il punto fisso è una contrazione in distanza di Kantorovich. Questo è il cuore della teoria del controllo stocastico, dell'apprendimento per rinforzo, della semantica quantitativa dei programmi probabilistici. La probabilità non è più una misura statica, ma una dinamica contrattiva su uno spazio metrico.

In questa prospettiva, la probabilità, la logica, la metrica e la coalgebra non sono più discipline separate, ma aspetti di un'unica struttura concettuale. La probabilità è una trasformazione, la logica è una metrica, la coalgebra è una sintassi, la distanza di Kantorovich è una geometria, il μ -calcolo è una teoria dei punti fissi. Il Volume III mostra che la probabilità non è solo una misura, ma una geometria del comportamento; non è solo una logica, ma una metrica della distinzione; non è solo una dinamica, ma una contrazione; non è solo una teoria della verità, ma una teoria della distanza.

Dinamiche epistemiche probabilistiche e logiche del cambiamento della conoscenza

La probabilità, quando viene interpretata non come misura statica ma come processo di aggiornamento, diventa una dinamica epistémica: un'evoluzione della conoscenza nel tempo. In questa prospettiva, la logica probabilistica non è più un linguaggio per descrivere proprietà di modelli, ma un linguaggio per descrivere trasformazioni della credenza. Ogni osservazione modifica la distribuzione di probabilità sui modelli; ogni azione modifica la struttura delle possibilità; ogni comunicazione modifica lo stato epistémico degli agenti. La probabilità diventa così una semantica del cambiamento della conoscenza, e la logica probabilistica diventa una logica dinamica epistémica.

Consideriamo un agente che ha una distribuzione μ sui modelli M di una teoria. Ogni formula φ definisce un sottoinsieme $[\varphi]$ dei modelli, e l'aggiornamento bayesiano $\mu \mapsto \mu|\varphi$ è una trasformazione della credenza. Ma questa trasformazione non è un semplice condizionamento: è un cambiamento epistémico che modifica la struttura delle possibilità. La logica dinamica epistémica classica descrive il cambiamento della conoscenza attraverso operatori modali come $[\varphi]$, che rappresentano l'atto di annunciare φ . Nella versione probabilistica, l'annuncio non è più un'operazione deterministica, ma

una trasformazione stocastica: l'agente non apprende ϕ con certezza, ma riceve un'evidenza che modifica la sua distribuzione. L'atto epistemico non è più un taglio netto, ma una contrazione informazionale.

La dinamica epistemica probabilistica è descritta da operatori di aggiornamento che agiscono su distribuzioni. Un annuncio incerto è una mappa $\mu \mapsto \alpha \mu|_{\phi} + (1-\alpha) \mu$, dove α è la forza dell'evidenza. Un'osservazione rumorosa è una mappa $\mu \mapsto \sum_i p_i \mu|_{\phi_i}$, dove p_i sono le probabilità dell'osservazione. Un'azione epistemica è una trasformazione $\mu \mapsto T(\mu)$, dove T è un operatore stocastico che rappresenta l'effetto dell'azione sulla credenza. La logica probabilistica dinamica è la sintassi naturale per descrivere tali operatori: formule come $[T]\phi$ significano che, dopo l'aggiornamento T , ϕ è altamente probabile.

Questa struttura permette di reinterpretare la conoscenza come un processo di contrazione della misura. Ogni osservazione riduce l'entropia della distribuzione, e quindi aumenta la conoscenza. La conoscenza non è più un insieme di modelli compatibili, ma una distribuzione concentrata. La certezza è il limite in cui la distribuzione collassa su un singolo modello. La logica epistemica classica è un caso limite della logica epistemica probabilistica: quando la distribuzione è degenerata, la conoscenza è classica; quando è diffusa, la conoscenza è probabilistica.

La dinamica epistemica probabilistica ci lascia modellare la comunicazione tra agenti. Ogni agente ha una distribuzione μ_i sui modelli; un annuncio pubblico modifica simultaneamente tutte le distribuzioni; un annuncio privato modifica solo alcune. La comunicazione è una trasformazione congiunta delle credenze, e la logica probabilistica dinamica è la sintassi naturale per descriverla. La nozione di conoscenza comune, fondamentale nella logica epistemica classica, diventa una proprietà della convergenza delle distribuzioni: un fatto è conoscenza comune se tutte le distribuzioni degli agenti assegnano probabilità 1 a quel fatto, e tutte assegnano probabilità 1 al fatto che tutte assegnano probabilità 1, e così via. La conoscenza comune è un punto fisso epistemico.

La dinamica epistemica probabilistica permette anche di modellare l'apprendimento. Un agente osserva una sequenza di evidenze e aggiorna la sua distribuzione. La convergenza della distribuzione verso la verità è garantita da teoremi di apprendimento bayesiano: se la verità è nel supporto iniziale, la probabilità assegnata al modello vero tende a 1 quasi sicuramente. Questo è il teorema di Doob, che mostra che l'apprendimento bayesiano è coerente: la credenza converge alla verità. La logica probabilistica dinamica è la sintassi naturale per descrivere questa convergenza: formule come $\mu_n(\phi) \rightarrow 1$ esprimono che la credenza in ϕ cresce nel tempo.

La dinamica epistemica probabilistica tratta, inoltre, l'inganno, la disinformazione, la manipolazione. Un agente può ricevere evidenze false, e la sua distribuzione può divergere dalla verità. La logica probabilistica dinamica può esprimere condizioni di robustezza epistemica: una teoria è robusta se la distribuzione converge alla verità nonostante perturbazioni. La robustezza è una proprietà metrica: la distanza tra la distribuzione dell'agente e la distribuzione vera deve rimanere piccola. La logica probabilistica dinamica è quindi una logica della resilienza epistemica.

In questa prospettiva, la probabilità non è più una misura statica, ma una dinamica della conoscenza; la logica non è più una teoria della verità, ma una teoria del cambiamento della verità epistemica; la semantica non è più una relazione tra formule e modelli, ma una relazione tra formule e trasformazioni della credenza. Il Volume III mostra che la probabilità è una teoria della conoscenza in movimento, e la logica probabilistica è la grammatica di questo movimento.

Probabilità, causalità e modelli strutturali

La distinzione tra correlazione e causalità è uno dei punti più profondi e più delicati dell'intera teoria della probabilità. La probabilità osservativa, per quanto raffinata, rimane confinata all'interno di una geometria simmetrica: essa registra che due variabili X e Y variano insieme, ma non può dire se X influenzi Y , se Y influenzi X , o se entrambe siano influenzate da una terza variabile Z . La quantità $\mu(Y|X)$ è cieca alla direzione della dipendenza, perché il condizionamento osservativo non rimuove le influenze che normalmente determinano X . Per distinguere tra correlazione e causalità occorre introdurre una struttura più ricca, che non si limiti a descrivere distribuzioni, ma che descriva meccanismi. Questa struttura è il modello causale strutturale.

In un modello causale, ogni variabile endogena V è determinata da una funzione strutturale f_V che prende come argomenti i valori dei genitori causali $Pa(V)$ e un rumore esogeno U_V . La forma generale dell'equazione è $V = f_V(Pa(V), U_V)$. L'insieme di tutte queste equazioni definisce un sistema strutturale F , e la distribuzione osservativa $\mu(V_1, \dots, V_n)$ è indotta dalla distribuzione dei rumori esogeni. Il grafo diretto aciclico associato al sistema, in cui un arco $X \rightarrow Y$ indica che X compare tra i genitori di Y , codifica la struttura causale. La probabilità osservativa è quindi un'ombra della struttura causale: essa è ottenuta integrando i rumori attraverso le funzioni strutturali.

La nozione di intervento è ciò che distingue la causalità dalla correlazione. Intervenire su X imponendo $X=x$ significa sostituire l'equazione $X = f_X(Pa(X), U_X)$ con la costante $X := x$. Questa operazione, denotata $do(X=x)$, rimuove tutte le influenze che normalmente determinano X . La nuova distribuzione $\mu^{\{do(X=x)\}}$ è ottenuta propagando questa modifica attraverso il sistema. La probabilità causale $\mu(Y=y | do(X=x))$ è quindi definita rispetto a un modello modificato, non rispetto alla distribuzione osservativa originale. È proprio questa differenza che permette di distinguere tra correlazione e causalità.

Per esempio, se Z causa sia X che Y , allora si può avere $\mu(Y|X) \neq \mu(Y)$, ma $\mu(Y|do(X)) = \mu(Y)$, perché intervenire su X non modifica il meccanismo che collega Z a Y . La differenza tra $\mu(Y|X)$ e $\mu(Y|do(X))$ è la differenza tra previsione e spiegazione. La probabilità osservativa è una teoria della previsione; la probabilità causale è una teoria della spiegazione.

Il calcolo causale di Pearl consente di esprimere $\mu(Y|do(X=x))$ in termini di quantità osservabili quando certe condizioni di separazione nel grafo sono soddisfatte. La formula più nota è la regola del back-door, che afferma che se Z blocca tutti i percorsi di back-door da X a Y , allora

$$\mu(Y | do(X=x)) = \sum_z \mu(Y | X=x, Z=z) \mu(Z=z).$$

Questa formula non è una tautologia probabilistica, ma una conseguenza della struttura causale. Essa mostra come la causalità introduca una logica aggiuntiva: non basta manipolare probabilità, occorre manipolare strutture.

La definizione di causa secondo Halpern e Pearl è controfattuale. $X=x$ è una causa di $Y=y$ se esiste un contesto u dei rumori esogeni e un insieme W di variabili tale che: nel contesto u , $X(u)=x$ e $Y(u)=y$; esiste un valore x' tale che, sostituendo $X:=x'$ e mantenendo W fissate a certi valori w , si ottiene $Y \neq y$; e nessun sottoinsieme più piccolo di W ha questa proprietà. Questa definizione formalizza l'idea che X sia una causa di Y se, tenendo fermo il resto del mondo quanto più possibile, cambiare X cambia Y . La responsabilità causale è definita come $1/(1+|W|)$, dove W è il più piccolo insieme di variabili che devono essere modificate insieme a X per alterare Y . Questa quantità misura quanto X sia "necessario" per Y .

La probabilità causale permette anche di definire effetti medi. L'effetto medio di X su Y è $E[Y | \text{do}(X=1)] - E[Y | \text{do}(X=0)]$, che è diverso da $E[Y | X=1] - E[Y | X=0]$ quando c'è confondimento. La differenza tra queste due quantità è la differenza tra previsione e spiegazione.

Ma la profondità della causalità emerge pienamente solo quando si considera la struttura controfattuale nel suo insieme. Ogni modello causale definisce una famiglia di mondi possibili, ciascuno corrispondente a un diverso intervento. Il mondo reale è quello determinato dal sistema F ; il mondo controfattuale in cui X è forzato a x è determinato dal sistema $F^{\wedge}x$; il mondo controfattuale in cui X è forzato a x e W è forzato a w è determinato dal sistema $F^{\wedge}\{x,w\}$. La semantica controfattuale è quindi definita non da una relazione di vicinanza arbitraria tra mondi, ma da una trasformazione strutturale del sistema. Questa è la differenza fondamentale tra la logica controfattuale di Lewis e la logica causale di Pearl: nella prima, la vicinanza tra mondi è un assioma; nella seconda, è una conseguenza della struttura causale.

La causalità introduce anche una nozione di indipendenza più fine di quella probabilistica. Due variabili X e Y possono essere indipendenti in senso osservativo, cioè $\mu(X,Y) = \mu(X)\mu(Y)$, ma dipendenti in senso causale, cioè $\mu(Y|\text{do}(X)) \neq \mu(Y)$. Questo accade quando X e Y sono collegati da percorsi che si attivano solo sotto intervento. Viceversa, due variabili possono essere dipendenti in senso osservativo ma indipendenti in senso causale, come nel caso del confondimento. La causalità introduce quindi una geometria più ricca della probabilità: una geometria direzionale, in cui la dipendenza ha una direzione e una forza.

La causalità permette anche di definire una nozione di spiegazione. Una spiegazione di $Y=y$ è un insieme di variabili $X_1=x_1, \dots, X_k=x_k$ tale che, nel modello causale, forzare queste variabili ai loro valori osservati rende $Y=y$ inevitabile. La spiegazione è quindi una proprietà controfattuale: $X_1=x_1, \dots, X_k=x_k$ spiegano $Y=y$ se, per ogni contesto u compatibile con l'osservazione, si ha $Y^{\wedge}\{x_1, \dots, x_k\}(u) = y$. La probabilità osservativa non può esprimere questa nozione, perché non distingue tra cause e conseguenze.

La causalità porta con sé una nozione di responsabilità. La responsabilità di $X=x$ per $Y=y$ è definita come l'inverso della dimensione minima del set W che deve essere modificato insieme a X per alterare Y . Questa quantità misura quanto X sia "necessario" per Y . La responsabilità è quindi una misura quantitativa della causalità, che integra la struttura controfattuale con la probabilità.

Infine, la causalità permette di unificare apprendimento e intervento. Un agente che apprende non osserva solo correlazioni, ma costruisce un modello strutturale del mondo. La probabilità osservativa è insufficiente per apprendere la causalità; servono interventi, esperimenti, manipolazioni. La logica probabilistica causale è la grammatica di questo apprendimento attivo: ogni intervento è un atto epistemico che modifica la struttura del modello. La causalità è quindi una teoria della conoscenza attiva, non passiva.

In questa prospettiva, la probabilità non è più una misura statica, ma una dinamica della conoscenza; la logica non è più una teoria della verità, ma una teoria del cambiamento della verità epistemica; la semantica non è più una relazione tra formule e modelli, ma una relazione tra formule e trasformazioni della credenza. La causalità è la struttura che permette di passare dalla descrizione alla spiegazione, dalla correlazione all'intervento, dalla previsione alla manipolazione.

Probabilità e continuità: processi stocastici, moto browniano e SDE

La probabilità continua rappresenta uno dei punti di svolta più radicali nella storia della matematica moderna. Nel passaggio dal discreto al continuo, la probabilità cambia natura: non è più una misura

su insiemi finiti o numerabili, ma una misura su spazi di funzioni, su traiettorie, su cammini. Il moto browniano, introdotto da Wiener, è il prototipo di questa trasformazione concettuale. Esso è definito da tre proprietà: $B_0 = 0$, gli incrementi $B_t - B_s$ sono indipendenti e distribuiti come una variabile normale $N(0, t-s)$, e la funzione $t \mapsto B_t$ è continua quasi sicuramente. Questa combinazione di continuità e irregolarità è la firma della probabilità continua: il moto browniano è continuo ovunque ma non è derivabile in nessun punto, quasi sicuramente. La probabilità continua è quindi la teoria matematica dell'irregolarità controllata.

La misura di Wiener μ_W è una misura su $C[0, \infty)$, lo spazio delle funzioni continue, e ogni evento naturale è un sottoinsieme di questo spazio. La probabilità di un evento come "il processo rimane nell'intervallo $[a, b]$ fino al tempo T " è $\mu_W(\{\omega : \omega(t) \in [a, b] \text{ per ogni } t \leq T\})$. La logica probabilistica continua è quindi una logica dei cammini: ogni formula descrive una proprietà di traiettorie, e la probabilità è una misura su insiemi di funzioni.

La presenza di un rumore continuo come B_t rende naturale considerare equazioni differenziali stocastiche (SDE). Una SDE della forma

$$dX_t = b(X_t, t) dt + \sigma(X_t, t) dB_t$$

è interpretata come l'equazione integrale

$$X_t = X_0 + \int_0^t b(X_s, s) ds + \int_0^t \sigma(X_s, s) dB_s.$$

L'integrale stocastico di Itô $\int \sigma dB$ è definito come limite in L^2 di somme del tipo $\sigma(X_{\{t_i\}}, t_i)(B_{\{t_{i+1}\}} - B_{\{t_i\}})$. La caratteristica fondamentale dell'integrale di Itô è che il termine quadratico non è trascurabile: $(dB_t)^2 = dt$, mentre $dt dB_t = 0$ e $(dt)^2 = 0$. Questa struttura infinitesimale è ciò che distingue la probabilità continua dalla probabilità classica: il rumore non è un'aggiunta esterna, ma una componente intrinseca della dinamica.

La formula di Itô per una funzione f sufficientemente regolare è

$$df(X_t) = f'(X_t) dX_t + (1/2) f''(X_t) \sigma(X_t, t)^2 dt.$$

Questa formula sostituisce la regola della catena classica e rappresenta la regola di derivazione fondamentale della logica probabilistica continua. Essa permette di calcolare l'evoluzione dell'aspettativa $E[f(X_t)]$ e quindi di dedurre proprietà quantitative del processo. La formula di Itô è la chiave per comprendere come la probabilità si propaga nel tempo in presenza di rumore.

La connessione con le equazioni alle derivate parziali emerge dal teorema di Feynman-Kac. Se consideriamo una funzione

$$u(x, t) = E[g(X_T) | X_t = x],$$

allora u soddisfa l'equazione parabolica

$$\partial_t u + b(x, t) \partial_x u + (1/2) \sigma(x, t)^2 \partial_{xx} u = 0$$

con condizione al bordo $u(x, T) = g(x)$. La probabilità continua è quindi una teoria delle soluzioni di PDE viste come aspettative di funzionali stocastici. La dinamica stocastica e la dinamica differenziale sono due facce della stessa struttura: una descrive l'evoluzione delle traiettorie, l'altra l'evoluzione delle densità.

La nozione di filtrazione $\{\mathcal{F}_t\}$ formalizza l'informazione disponibile fino al tempo t . Un processo X_t è adattato se X_t è $\mathcal{F}_t$ -misurabile. Una martingala è un processo tale che $E[X_t | \mathcal{F}_s] = X_s$ per $s \leq t$. Le martingale sono invarianti epistemici: rappresentano quantità che non cambiano in aspettativa quando si aggiorna l'informazione. Il teorema di arresto (optional stopping) afferma che, sotto condizioni tecniche, se τ è un tempo di arresto, allora $E[X_\tau] = E[X_0]$ per ogni martingala X_t . Questa proprietà è la base logica di molti argomenti probabilistici continui: hitting times, barriere, stopping rules.

La probabilità continua è quindi una logica dei cammini, delle filtrazioni, delle martingale, delle SDE, delle PDE associate. È una logica infinitesimale, in cui le formule descrivono proprietà di traiettorie e le probabilità sono misure su spazi funzionali. È anche una logica termica: la diffusione della probabilità è descritta dall'equazione di Fokker-Planck

$$\partial_t p = -\partial_x (b p) + (1/2) \partial_{xx} (\sigma^2 p),$$

che è la versione densitaria della dinamica stocastica. La probabilità non è più statica, ma fluisce come un fluido in uno spazio di stati.

La continuità si riallaccia ad una nozione di tempo che non è più discreto ma infinitesimale. La probabilità continua è una teoria del tempo infinitesimale: ogni istante porta nuova informazione, e la probabilità si aggiorna in modo continuo. La nozione di “quasi ovunque” diventa più sottile: eventi di probabilità zero possono essere inevitabili, come il fatto che il moto browniano attraversi ogni intervallo infinitesimale.

La probabilità continua è anche una teoria della regolarità e dell'irregolarità. Il moto browniano è continuo ma non derivabile; le soluzioni delle SDE sono spesso Hölder-continue ma non più regolari; le densità delle soluzioni delle PDE associate sono lisce anche quando le traiettorie non lo sono. La probabilità continua è quindi una teoria della dualità tra regolarità e irregolarità: le traiettorie sono irregolari, ma le densità sono regolari.

Infine, la probabilità continua è il ponte tra probabilità classica e probabilità quantistica. Il cammino di un'ampiezza quantistica può essere visto, in certe formulazioni path-integral, come un integrale su traiettorie in cui a ogni cammino è associata una fase complessa. Il moto browniano è la controparte reale di questa costruzione: un integrale su traiettorie con pesi reali, non complessi. In questa analogia, la logica probabilistica continua è la sorella reale della logica delle ampiezze quantistiche: entrambe sono logiche su spazi di cammini, ma in una i pesi sono probabilità, nell'altra sono ampiezze.

La probabilità continua è quindi una teoria della continuità irregolare, una teoria delle traiettorie, una teoria delle densità, una teoria delle martingale, una teoria delle SDE, una teoria delle PDE, una teoria dell'informazione infinitesimale. È una delle strutture più profonde dell'intera matematica moderna.

Quadri Argomentativi Probabilistici

La teoria dell'argomentazione formale nasce come tentativo di modellare il ragionamento non-monotono attraverso strutture astratte di attacco e difesa. Nel quadro di Dung, un sistema argomentativo è costituito da un insieme di argomenti A e da una relazione di attacco $\rightarrow \subseteq A \times A$. La semantica classica assegna a ogni insieme $S \subseteq A$ uno status di accettabilità basato su criteri come completezza, stabilità, preferibilità. Tuttavia, questa struttura è intrinsecamente qualitativa: un attacco è presente o assente, un argomento è accettato o rifiutato. Per modellare l'incertezza, occorre introdurre una struttura probabilistica che permetta di assegnare pesi, gradi di credenza, distribuzioni

di probabilità sugli attacchi, sugli argomenti o sugli stati di accettazione. Nasce così il quadro argomentativo probabilistico, in cui la relazione di attacco è arricchita da una funzione μ che assegna a ogni coppia (a,b) un valore $\mu(a \rightarrow b) \in [0,1]$, interpretato come la probabilità che l'attacco sia effettivo.

L'idea fondamentale è che la struttura argomentativa non è più deterministica, ma aleatoria: un attacco può essere presente con probabilità $\mu(a \rightarrow b)$, e l'intero grafo degli attacchi diventa una variabile aleatoria che assume valori in $2^{A \times A}$. Ogni realizzazione ω del grafo determina un quadro di Dung classico, e la probabilità che un argomento sia accettato sotto una certa semantica σ è definita come $\mu(\{\omega : a \text{ è accettato in } \omega \text{ secondo } \sigma\})$. La funzione di accettazione probabilistica $\alpha_\sigma(a) = \mu(a \text{ accettato secondo } \sigma)$ è quindi una generalizzazione quantitativa della semantica classica. Essa permette di distinguere tra argomenti robusti, che rimangono accettati nella maggior parte delle realizzazioni, e argomenti fragili, che dipendono da configurazioni rare del grafo.

La presenza di attacchi probabilistici introduce una nozione di difesa probabilistica. Un argomento a difende b con probabilità p se la probabilità che a attacchi tutti gli attaccanti di b è almeno p . Questa nozione è formalizzata definendo la probabilità che un insieme $S \subseteq A$ difenda b come $\mu(\{\omega : \text{per ogni } c \text{ che attacca } b \text{ in } \omega, \text{ esiste un } d \in S \text{ che attacca } c \text{ in } \omega\})$. La difesa non è più un fatto binario, ma una proprietà graduata. La semantica probabilistica dell'argomentazione deve quindi essere definita come un punto fisso di un operatore che assegna a ogni argomento un valore in $[0,1]$, rappresentante il grado di accettazione. Una possibile definizione è $\alpha(a) = 1 - \sup_{b \text{ attacca } a} \mu(b \text{ accettato})$, che generalizza la semantica stabile. Un'altra è $\alpha(a) = \inf_{b \text{ attacca } a} (1 - \alpha(b))$, che generalizza la semantica completa. In entrambi i casi, α è un punto fisso di un operatore monotono su $[0,1]^A$, e il teorema di Knaster–Tarski garantisce l'esistenza di punti fissi minimi e massimi.

La dinamica probabilistica dell'argomentazione emerge quando si considera l'evoluzione nel tempo della funzione μ . Gli attacchi possono essere appresi, dimenticati, rafforzati o indeboliti sulla base di nuove evidenze. La probabilità $\mu_t(a \rightarrow b)$ al tempo t può essere aggiornata secondo una regola bayesiana $\mu_{t+1}(a \rightarrow b) = \mu_t(a \rightarrow b \mid \text{evidenza}_t)$, oppure secondo una dinamica stocastica come $\mu_{t+1}(a \rightarrow b) = \mu_t(a \rightarrow b) + \eta_t \Delta_t(a \rightarrow b)$, dove Δ_t è un segnale di rinforzo e η_t è un tasso di apprendimento. La funzione di accettazione $\alpha_t(a)$ diventa quindi una traiettoria in $[0,1]^A$, e la semantica dell'argomentazione diventa una dinamica. La probabilità non è più solo una misura statica, ma una trasformazione che evolve nel tempo.

La presenza di attacchi probabilistici permette anche di definire una nozione di coerenza probabilistica. Un insieme $S \subseteq A$ è coerente con probabilità p se $\mu(\{\omega : S \text{ è privo di conflitti in } \omega\}) \geq p$. Questa nozione permette di distinguere tra insiemi coerenti in senso forte ($p=1$) e insiemi coerenti in senso debole ($p<1$). La semantica probabilistica può quindi essere definita come la ricerca di insiemi S che massimizzano la coerenza probabilistica. Questa prospettiva collega l'argomentazione probabilistica alla teoria dei giochi, alla teoria della decisione e alla teoria dell'evidenza.

La probabilità è incentrata sull'incertezza degli argomenti stessi. Ogni argomento a può essere presente con probabilità $\mu(a)$, e la struttura argomentativa diventa un oggetto aleatorio in cui sia gli argomenti sia gli attacchi sono incerti. La funzione di accettazione probabilistica deve quindi tener conto sia dell'incertezza strutturale sia dell'incertezza relazionale. La semantica risultante è una funzione $\alpha: A \rightarrow [0,1]$ che assegna a ogni argomento un grado di accettazione che integra tutte le fonti di incertezza.

Infine, la probabilità permette di definire una nozione di forza argomentativa. La forza di un argomento a può essere definita come la probabilità che a sia accettato sotto tutte le semantiche classiche, cioè $\alpha(a) = \mu(\{\omega : a \text{ è accettato in } \omega \text{ secondo ogni semantica } \sigma\})$. Questa definizione

generalizza la nozione di argomento universalmente accettabile. La probabilità permette quindi di misurare la robustezza argomentativa.

Il quadro argomentativo probabilistico è quindi una generalizzazione quantitativa del quadro di Dung, in cui la relazione di attacco è arricchita da una struttura probabilistica e la semantica è definita come un punto fisso di un operatore su $[0,1]^A$. La probabilità non è un'aggiunta esterna, ma una componente intrinseca della struttura argomentativa. Essa permette di modellare l'incertezza, la dinamica, la robustezza, la forza e la coerenza in modo naturale e rigoroso.

Probabilità infinitesimali nella logica

La probabilità infinitesimale nasce dall'esigenza di trattare eventi di probabilità zero come oggetti dotati di significato logico, semantico e inferenziale. Nella teoria classica della misura, un evento E con $\mu(E)=0$ è un evento che può essere ignorato in ogni calcolo integrale, ma non può essere condizionato: l'espressione $\mu(A|E)$ è indefinita. Tuttavia, molte strutture logiche richiedono di ragionare proprio su eventi di misura zero, come nel caso dei condizionamenti iterati, delle probabilità controfattuali, delle probabilità epistemiche su ipotesi infinitamente specifiche, o delle probabilità di sequenze infinite in cui ogni singola sequenza ha probabilità zero. Per superare questo limite, occorre introdurre una nozione di probabilità che ammetta valori infinitesimali, cioè quantità ε tali che $0 < \varepsilon < r$ per ogni reale positivo r . L'analisi non-standard fornisce un quadro rigoroso per trattare tali quantità.

Nell'approccio non-standard, si considera un'estensione ${}^*\mathbb{R}$ dei numeri reali che contiene numeri infinitamente piccoli e infinitamente grandi. Una probabilità μ definita su uno spazio di eventi può essere estesa a una probabilità ${}^*\mu$ che assume valori in ${}^*\mathbb{R}$. Un evento E può quindi avere probabilità $\mu(E)=\varepsilon$, dove ε è un infinitesimale. La probabilità standard $\mu(E)$ è ottenuta come $st(\mu(E))$, cioè come parte standard di ${}^*\mu(E)$. La misura di Loeb permette di trasformare una misura interna ${}^*\mu$ in una misura standard μ_L definita su una σ -algebra generata dagli insiemi interni. Questa costruzione permette di trattare eventi infinitesimali come eventi di misura zero dotati di struttura interna. La probabilità infinitesimale non è quindi una violazione della teoria classica, ma un suo raffinamento: essa permette di distinguere tra eventi di misura zero che sono “veramente impossibili” e eventi di misura zero che sono “infinitamente improbabili”.

La nozione di probabilità condizionata regolare emerge proprio da questa distinzione. Una probabilità condizionata regolare è una funzione $\mu(A|B)$ definita per quasi ogni B , tale che $\mu(A \cap B) = \int_B \mu(A|B) d\mu$. Tuttavia, questa definizione non assegna un valore a $\mu(A|B)$ quando $\mu(B)=0$. L'analisi non-standard permette di definire $\mu(A|B)$ anche quando $\mu(B)=0$, interpretando $\mu(A \cap B)/\mu(B)$ come ${}^*\mu(A \cap B)/{}^*\mu(B)$, che può essere un numero finito anche quando ${}^*\mu(B)$ è infinitesimale. La parte standard di questo rapporto definisce $\mu(A|B)$ anche quando $\mu(B)=0$. La probabilità infinitesimale permette quindi di definire il condizionamento su eventi di misura zero in modo rigoroso.

Le funzioni di Popper rappresentano un altro approccio alla probabilità condizionale su eventi di misura zero. Una funzione di Popper è una funzione $\mu(\phi|\psi)$ definita per tutte le coppie di formule ϕ , ψ , anche quando $\mu(\psi)=0$, che soddisfa le proprietà assiomatiche della probabilità condizionale. La definizione fondamentale è che $\mu(\phi \wedge \psi) = \mu(\phi|\psi) \mu(\psi)$, ma $\mu(\psi)$ può essere zero. In questo caso, $\mu(\phi|\psi)$ è definita come un limite di $\mu(\phi|\psi_n)$, dove ψ_n è una sequenza di eventi che approssimano ψ dall'interno. La funzione di Popper è quindi una generalizzazione della probabilità condizionale che permette di trattare eventi di misura zero come condizioni legittime. Essa è particolarmente utile nella logica condizionale, dove formule come $\phi \rightarrow \psi$ sono interpretate come $\mu(\psi|\phi)$ elevata a un certo livello di soglia.

La probabilità infinitesimale è in grado di definire una nozione di probabilità a priori su spazi infiniti. In un modello bayesiano classico, una distribuzione a priori su $\mathbb{N}$ deve assegnare probabilità positive a ogni numero naturale, ma la somma di infiniti numeri positivi non può essere 1. La probabilità infinitesimale permette di assegnare a ogni n una probabilità ε , dove ε è un infinitesimale, e la somma $\sum_{n=1}^{\infty} \varepsilon$ è un numero finito la cui parte standard è 1. La probabilità infinitesimale permette quindi di definire distribuzioni uniformi su insiemi infiniti, che sono impossibili nella teoria classica.

La probabilità infinitesimale è anche fondamentale nella teoria dei processi stocastici. Nel moto browniano, ogni traiettoria ha probabilità zero, ma l'insieme delle traiettorie che soddisfano una certa proprietà può avere probabilità positiva. La probabilità infinitesimale permette di interpretare la probabilità di una singola traiettoria come un infinitesimale ε , e la probabilità di un insieme di traiettorie come la somma di infiniti infinitesimali. La misura di Loeb permette di trasformare questa somma in una misura standard. La probabilità infinitesimale è quindi una teoria della granularità infinitesimale della probabilità continua.

La probabilità infinitesimale è altresì importante nella teoria della decisione. In molte situazioni, un agente deve considerare eventi di probabilità zero come possibili, perché la loro esclusione porterebbe a decisioni irrazionali. La probabilità infinitesimale permette di assegnare a tali eventi una probabilità ε , che è sufficiente per influenzare la decisione senza distorcere la distribuzione complessiva. La teoria della decisione infinitesimale è quindi una teoria della razionalità in presenza di eventi infinitamente improbabili.

Infine, la probabilità infinitesimale è fondamentale nella logica epistemica. Un agente può considerare un evento φ come possibile anche quando $\mu(\varphi)=0$, perché la sua conoscenza è limitata. La probabilità infinitesimale permette di modellare questa possibilità assegnando a φ una probabilità ε . La logica epistemica infinitesimale è quindi una teoria della conoscenza imperfetta in presenza di eventi infinitamente improbabili.

La probabilità infinitesimale è quindi una generalizzazione della probabilità classica che permette di trattare eventi di misura zero come oggetti dotati di significato logico, semantico e inferenziale. Essa integra l'analisi non-standard, le funzioni di Popper, le probabilità condizionate regolari, le misure di Loeb, la teoria della decisione e la logica epistemica in un'unica struttura coerente. La probabilità infinitesimale non è un'alternativa alla probabilità classica, ma un suo raffinamento: essa permette di trattare la probabilità con una granularità infinitesimale che la teoria classica non può esprimere.

Probabilità nella teoria dei topoi

La probabilità nella teoria dei topoi rappresenta uno dei punti più alti della generalizzazione concettuale della misura. Nella teoria classica, una probabilità μ è una misura definita su una σ -algebra di sottoinsiemi di uno spazio X , e una variabile aleatoria è una funzione misurabile $X \rightarrow \mathbb{R}$. Questa struttura, pur essendo estremamente potente, è legata alla logica classica e alla teoria degli insiemi. La teoria dei topoi permette di riformulare la probabilità in un contesto più generale, in cui gli insiemi sono sostituiti da oggetti di un topos, le funzioni da morfismi, e la logica classica da una logica interna intuizionista. In questo contesto, la probabilità non è più una misura su insiemi, ma un morfismo $p : X \rightarrow [0,1]$, dove $[0,1]$ è un oggetto del topos che rappresenta l'intervallo unitario. Questa generalizzazione, introdotta da Lawvere, permette di interpretare la probabilità come una logica suboggettiva: un predicato φ su X è interpretato come un morfismo $\chi\varphi : X \rightarrow \Omega$, dove Ω è l'oggetto verità del topos, e la probabilità di φ è un morfismo $\mu(\varphi) : 1 \rightarrow [0,1]$ ottenuto componendo $\chi\varphi$ con p .

La monade di Giraud rappresenta la probabilità come un costrutto categoriale. Nel contesto classico, la monade di Giraud G associa a ogni spazio misurabile X lo spazio $G(X)$ delle misure di probabilità su

X , e a ogni funzione misurabile $f : X \rightarrow Y$ il morfismo $G(f) : G(X) \rightarrow G(Y)$ definito da $G(f)(\mu) = \mu \circ f^{-1}$. Questa struttura cattura la natura functoriale della probabilità: una trasformazione di spazi induce una trasformazione delle misure. Nel contesto di un topos, la monade di Girard può essere definita internamente, sostituendo gli spazi misurabili con oggetti del topos e le funzioni misurabili con morfismi. Una misura interna su X è un morfismo $\mu : X \rightarrow [0,1]$ che soddisfa le proprietà della misura nella logica interna. La monade di Girard interna G è definita come un endofunctor $G : \mathbb{U} \rightarrow \mathbb{U}$, dove $\mathbb{U}$ è il topos, che associa a X l'oggetto $G(X)$ delle misure interne su X . La struttura monadica (η, μ) permette di definire integrazione, pushforward e trasformazioni di misura nella logica interna.

Una variabile aleatoria interna è un morfismo $\xi : 1 \rightarrow X$, interpretato come un elemento generalizzato di X . Nella logica interna, un elemento generalizzato non è un punto, ma una famiglia di punti parametrizzata da contesti. La probabilità di un predicato ϕ su X è definita come $\mu(\phi) = \int \chi_\phi d\mu$, dove l'integrale è definito internamente tramite la struttura monadica. Questa definizione permette di trattare variabili aleatorie che non hanno un valore puntuale, ma solo un comportamento interno. La probabilità categoriale è quindi una teoria della misura senza punti, in cui la nozione di punto è sostituita dalla nozione di elemento generalizzato.

La logica interna di un topos è intuizionista, e ciò ha conseguenze profonde sulla probabilità. Nella logica classica, un predicato ϕ è vero o falso; nella logica interna, ϕ può essere vero in alcuni contesti e indeterminato in altri. La probabilità $\mu(\phi)$ non è quindi una misura della frequenza di verità, ma una misura della forza con cui ϕ è verificato internamente. La probabilità categoriale è quindi una logica della verità graduata, in cui la probabilità misura la stabilità interna di un predicato. Questa prospettiva collega la probabilità alla logica fuzzy, alla logica modale e alla logica topologica.

La probabilità categoriale permette di trattare spazi di probabilità non classici, come gli spazi quantistici. In un topos costruito a partire da un'algebra di von Neumann, l'oggetto verità Ω non è un booleano, ma un reticolo ortomodulare. La probabilità $\mu : X \rightarrow [0,1]$ è definita come uno stato su un'algebra di proiezioni, e la logica interna è la logica quantistica. La probabilità categoriale unifica quindi la probabilità classica e la probabilità quantistica in un'unica struttura. La distinzione tra classico e quantistico non è più una distinzione tra due teorie, ma una distinzione tra due topoi.

La probabilità categoriale interpreta la probabilità come una teoria della trasformazione. Un kernel stocastico $K : X \rightarrow G(Y)$ è un morfismo che associa a ogni $x \in X$ una misura su Y . Nel contesto categoriale, un kernel è un morfismo $K : X \rightarrow G(Y)$ nella categoria $\mathbb{U}$. La composizione di kernel è definita tramite la struttura monadica, e la probabilità diventa una teoria della composizione di trasformazioni. Questa prospettiva è fondamentale nella teoria dei processi stocastici, nella teoria dei programmi probabilistici e nella teoria delle coalgebre probabilistiche.

La probabilità categoriale è quindi una generalizzazione profonda della probabilità classica. Essa sostituisce gli insiemi con oggetti di un topos, le funzioni con morfismi, le misure con morfismi verso $[0,1]$, le variabili aleatorie con elementi generalizzati, la logica classica con la logica interna. La probabilità non è più una misura su insiemi, ma una logica della verità graduata in un contesto categoriale. Essa unifica la probabilità classica, la probabilità quantistica, la teoria della misura, la teoria delle categorie e la logica intuizionista in un'unica struttura coerente.

Logica quantistica e probabilità

La probabilità quantistica nasce come risposta alla crisi della logica classica nel descrivere fenomeni fisici in cui la struttura degli eventi non è più distributiva. Nella teoria classica, gli eventi formano una σ -algebra booleana, e la probabilità μ è una misura additiva definita su questa struttura. Nella

teoria quantistica, gli eventi sono proiezioni in uno spazio di Hilbert H , e la loro struttura non è più un reticolo booleano, ma un reticolo ortomodulare. La probabilità μ non è più una misura su insiemi, ma uno stato ρ , cioè un operatore positivo di traccia uno, che assegna a ogni proiezione P una probabilità $\mu(P) = \text{Tr}(\rho P)$. Questa trasformazione concettuale è radicale: la probabilità non è più una misura su un insieme di mondi possibili, ma una valutazione su un insieme di osservabili incompatibili. La logica quantistica è quindi una logica della non-commutatività.

La struttura degli eventi quantistici è determinata dal reticolo delle proiezioni $\mathcal{P}(H)$. Due proiezioni P e Q sono compatibili se commutano, cioè $PQ = QP$; in questo caso, esiste un sottospazio comune che rappresenta la loro congiunzione. Se non commutano, la loro congiunzione non è definita come un'intersezione di insiemi, ma come la proiezione sul sottospazio generato dai vettori che appartengono simultaneamente ai due sottospazi in un senso generalizzato. La disgiunzione è definita come la proiezione sulla somma dei sottospazi. La negazione è definita come il complemento ortogonale. Tuttavia, la distributività fallisce: in generale, $P \wedge (Q \vee R) \neq (P \wedge Q) \vee (P \wedge R)$. Questo fallimento della distributività è la firma logica della meccanica quantistica. La logica quantistica non è una logica classica con probabilità strane, ma una logica diversa, in cui la struttura degli eventi è non-distributiva.

La probabilità quantistica è definita da uno stato ρ , che è un operatore positivo di traccia uno. La probabilità di un evento P è $\mu(P) = \text{Tr}(\rho P)$. Questa definizione generalizza la probabilità classica: se H è uno spazio di dimensione n e ρ è diagonale rispetto a una base ortonormale $\{e_i\}$, allora $\mu(P_i) = \rho_{ii}$ è la probabilità classica dell'evento i . Tuttavia, se ρ non è diagonale, la probabilità non è più una misura su un insieme di mondi possibili, ma una valutazione su un insieme di osservabili incompatibili. La probabilità quantistica è quindi una teoria della coerenza tra osservazioni incompatibili.

La non-commutatività degli eventi quantistici si manifesta nelle disuguaglianze di Bell. Consideriamo due osservatori che misurano osservabili A e B su un sistema bipartito. Se la probabilità fosse classica, le correlazioni tra A e B dovrebbero soddisfare certe disuguaglianze, come $|E(AB) + E(A'B) + E(AB') - E(A'B')| \leq 2$. Tuttavia, nella teoria quantistica, queste disuguaglianze possono essere violate: esistono stati ρ e osservabili A, A', B, B' tali che $|E(AB) + E(A'B) + E(AB') - E(A'B')|$ può raggiungere $2\sqrt{2}$. Questa violazione non è un'anomalia probabilistica, ma una conseguenza della struttura logica non-distributiva degli eventi quantistici. Le disuguaglianze di Bell sono quindi teoremi di non-classicità logica: esse mostrano che la logica degli eventi quantistici non può essere rappresentata come una σ -algebra booleana.

La logica quantistica è anche contestuale. Il teorema di Kochen–Specker mostra che non esiste una funzione v che assegni a ogni proiezione P un valore $v(P) \in \{0, 1\}$ tale che $v(P+Q) = v(P) + v(Q)$ quando P e Q sono ortogonali. In altre parole, non esiste una valutazione globale che assegni valori di verità agli eventi quantistici in modo coerente con la struttura del reticolo. La verità di un evento quantistico dipende dal contesto di misura, cioè dall'insieme di osservabili compatibili che vengono misurati insieme. La logica quantistica è quindi una logica contestuale: la verità non è una proprietà assoluta, ma una proprietà relativa al contesto.

La probabilità quantistica è quindi una teoria della coerenza contestuale. Uno stato ρ assegna a ogni proiezione P una probabilità $\mu(P) = \text{Tr}(\rho P)$, ma questa probabilità non può essere interpretata come la misura di un insieme di mondi possibili. Essa è una valutazione contestuale che dipende dalla struttura non-commutativa degli osservabili. La probabilità quantistica è quindi una teoria della coerenza tra osservazioni incompatibili.

La logica quantistica può essere vista come una generalizzazione della logica modale. In una logica modale, la verità di una formula dipende da un mondo possibile e da una relazione di accessibilità. Nella logica quantistica, la verità di un evento dipende da un contesto di misura e dalla relazione di compatibilità tra osservabili. La probabilità quantistica è quindi una logica modale non-classica, in cui la modalità è determinata dalla compatibilità degli osservabili. Questa prospettiva permette di interpretare la meccanica quantistica come una teoria modale della conoscenza fisica.

La probabilità quantistica è pure una teoria della trasformazione. Un'operazione quantistica è un'applicazione completamente positiva $T : \rho \mapsto \sum_i K_i \rho K_i^\dagger$, dove i K_i sono operatori di Kraus. Questa trasformazione generalizza il pushforward delle misure classiche. La probabilità quantistica è quindi una teoria della dinamica degli stati, in cui la trasformazione degli stati è determinata da operatori non-commutativi. La composizione delle operazioni quantistiche è associativa, ma non commutativa. La probabilità quantistica è quindi una teoria della trasformazione non-commutativa.

Infine, la probabilità quantistica è una teoria della misura senza punti. Nella teoria classica, una misura μ è definita su un insieme di punti. Nella teoria quantistica, non esistono punti: gli stati puri non sono punti, ma vettori di uno spazio di Hilbert modulo fase. La probabilità quantistica è quindi una teoria della misura su uno spazio senza punti, in cui la nozione di punto è sostituita dalla nozione di stato puro. Questa prospettiva collega la probabilità quantistica alla probabilità categoriale e alla teoria dei topoi.

La logica quantistica e la probabilità quantistica rappresentano quindi una generalizzazione profonda della logica e della probabilità classiche. Esse sostituiscono la σ -algebra booleana con un reticolo ortomodulare, la misura con uno stato ρ , la congiunzione con la proiezione sul sottospazio comune, la disgiunzione con la somma dei sottospazi, la negazione con il complemento ortogonale, la verità con la compatibilità contestuale. La probabilità non è più una misura su un insieme di mondi possibili, ma una valutazione contestuale su un insieme di osservabili incompatibili. La logica quantistica è quindi una logica della non-commutatività, della contestualità e della coerenza.

Logiche probabilistiche a punto fisso

La logica dei punti fissi probabilistici nasce dall'esigenza di descrivere proprietà che non si esauriscono in un singolo istante, ma che emergono come limiti, equilibri, comportamenti stazionari o condizioni ricorsive di sistemi dinamici. Nella logica classica, il μ -calcolo di Kozen fornisce un linguaggio potente per esprimere proprietà ricorsive tramite operatori di punto fisso μ e ν , che rappresentano rispettivamente il minimo e il massimo punto fisso di un operatore monotono. Nella logica probabilistica, questa struttura deve essere generalizzata per trattare operatori che agiscono non su insiemi, ma su funzioni che assegnano a ogni stato un valore in $[0,1]$, interpretato come un grado di soddisfacimento o una probabilità. La logica probabilistica a punto fisso è quindi una logica quantitativa, in cui le formule non sono più vere o false, ma assumono valori reali compresi tra 0 e 1.

La struttura fondamentale è un sistema di transizione probabilistico, descritto da un kernel stocastico K che assegna a ogni stato x una misura $K(x, \cdot)$ sugli stati successivi. Una formula ϕ della logica probabilistica è interpretata come una funzione $\llbracket \phi \rrbracket : X \rightarrow [0,1]$, e gli operatori logici sono interpretati come trasformazioni di tali funzioni. Per esempio, l'operatore modale probabilistico $\langle a \rangle \phi$ è interpretato come $\llbracket \langle a \rangle \phi \rrbracket(x) = \int \llbracket \phi \rrbracket(y) dK_a(x,y)$, dove K_a è il kernel associato all'azione a . Questa definizione generalizza la semantica modale classica: invece di chiedere se esiste un successore che soddisfa ϕ , si calcola la probabilità che un successore soddisfi ϕ . La logica probabilistica è quindi una logica dell'aspettativa.

Gli operatori di punto fisso μ e ν sono interpretati come il minimo e il massimo punto fisso di un operatore monotono $F : [0,1]^X \rightarrow [0,1]^X$. La monotonicità è garantita dalla struttura della logica: se $\varphi \leq \psi$ punto per punto, allora $\llbracket \langle a \rangle \varphi \rrbracket \leq \llbracket \langle a \rangle \psi \rrbracket$. Il teorema di Knaster–Tarski garantisce che ogni operatore monotono su un reticolo completo ha un punto fisso minimo e uno massimo. Nel contesto probabilistico, il reticolo è $[0,1]^X$, ordinato punto per punto, e i punti fissi sono funzioni $f : X \rightarrow [0,1]$ tali che $F(f)=f$. Il punto fisso minimo μF rappresenta la soluzione più bassa compatibile con la definizione ricorsiva, mentre il punto fisso massimo νF rappresenta la soluzione più alta. La logica probabilistica a punto fisso permette quindi di definire proprietà come “la probabilità di raggiungere un certo insieme di stati”, “la probabilità di soddisfare una proprietà ricorsiva”, “la probabilità di evitare un insieme di stati per sempre”.

Un esempio fondamentale è la probabilità di raggiungere un insieme di stati T . La funzione $f(x)=\mu_x(\Diamond T)$, cioè la probabilità che un processo iniziato in x raggiunga T , è il minimo punto fisso dell’operatore F definito da $F(g)(x)=1$ se $x \in T$, e $F(g)(x)=\int g(y) dK(x,y)$ altrimenti. Questa definizione è ricorsiva: la probabilità di raggiungere T da x è 1 se x è già in T , altrimenti è la media delle probabilità di raggiungere T dai successori di x . Il punto fisso minimo rappresenta la soluzione naturale di questa equazione. La logica probabilistica a punto fisso permette quindi di esprimere proprietà di raggiungibilità probabilistica.

Un altro esempio è la probabilità di evitare un insieme di stati T per sempre. La funzione $f(x)=\mu_x(\Box \neg T)$, cioè la probabilità che un processo iniziato in x non entri mai in T , è il massimo punto fisso dell’operatore F definito da $F(g)(x)=0$ se $x \in T$, e $F(g)(x)=\int g(y) dK(x,y)$ altrimenti. Questa definizione è duale alla precedente: la probabilità di evitare T è 0 se x è in T , altrimenti è la media delle probabilità di evitare T dai successori di x . Il punto fisso massimo rappresenta la soluzione naturale di questa equazione. La logica probabilistica a punto fisso permette quindi di esprimere proprietà di sicurezza probabilistica.

La logica probabilistica a punto fisso è una teoria delle distribuzioni stazionarie. Una distribuzione stazionaria π è una soluzione dell’equazione $\pi=\pi P$, dove P è la matrice di transizione di una catena di Markov. Questa equazione è un punto fisso: π è un punto fisso dell’operatore F definito da $F(\mu)=\mu P$. La monotonicità non è più garantita, ma la struttura lineare permette di applicare teoremi di punto fisso più generali. La logica probabilistica a punto fisso permette quindi di trattare le distribuzioni stazionarie come soluzioni di equazioni ricorsive.

La logica probabilistica a punto fisso può considerarsi una teoria della semantica dei programmi probabilistici. Un programma probabilistico può essere visto come un sistema di transizione probabilistico, e la sua semantica denotazionale è definita come un punto fisso di un operatore che descrive l’effetto del programma sulle distribuzioni. Per esempio, la semantica di un ciclo `while  $\varphi$  do  $P$` è definita come il minimo punto fisso dell’operatore $F(\mu)=\mu\varphi \circ \llbracket P \rrbracket + \mu\{\neg\varphi\}$, dove μ_φ è la restrizione della distribuzione agli stati che soddisfano φ . La logica probabilistica a punto fisso permette quindi di definire la semantica dei programmi probabilistici in modo rigoroso.

Infine, la logica probabilistica a punto fisso è una teoria della coinduzione quantitativa. Il punto fisso massimo νF rappresenta la soluzione più grande compatibile con una definizione ricorsiva, e può essere interpretato come una proprietà invariabile. La coinduzione probabilistica permette di dimostrare proprietà come “la probabilità di soddisfare una certa proprietà è almeno p ” mostrando che una funzione g è un prefisso di un punto fisso. La logica probabilistica a punto fisso è quindi una teoria della dimostrazione quantitativa.

La logica probabilistica a punto fisso unifica la teoria dei punti fissi, la teoria delle probabilità, la teoria dei processi stocastici, la teoria dei programmi probabilistici e la teoria della dimostrazione

quantitativa in un'unica struttura coerente. Essa permette di esprimere proprietà ricorsive probabilistiche, di definire distribuzioni stazionarie, di descrivere la semantica dei programmi probabilistici, di dimostrare proprietà invarianti e di trattare la probabilità come una logica quantitativa. La logica probabilistica a punto fisso è quindi una delle strutture più profonde della teoria moderna della probabilità e della logica.

Processi stocastici e misura sui cammini

La probabilità, nei primi due volumi, è stata trattata innanzitutto come struttura logica e semantica su spazi di stati e su modelli, e poi come trasformazione, come operatore che agisce su configurazioni, su credenze, su strutture argomentative, su oggetti categoriali e su logiche a punto fisso. Con il passaggio al terzo volume, la scena cambia in modo radicale: la probabilità non è più vista solo come valutazione su enunciati o come trasformazione discreta di distribuzioni, ma come teoria dei cammini, come misura su spazi di funzioni, come geometria di traiettorie e come analisi infinitesimale del caso. Qui la probabilità smette di essere soltanto una funzione μ che assegna numeri a insiemi di stati; diventa una misura su spazi come $C([0, T], \mathbb{R}^n)$, cioè sullo spazio delle funzioni continue da un intervallo di tempo $[0, T]$ in $\mathbb{R}^n$. Un “evento” non è più solo un sottoinsieme dello spazio degli stati, ma un sottoinsieme di cammini, di funzioni intere nel tempo. Il passaggio concettuale è netto: non misuriamo più “dove” si trova il sistema, ma “come si muove”.

Il concetto fondamentale è quello di processo stocastico. Un processo stocastico non è altro che una famiglia $\{X_t : t \in I\}$ di variabili aleatorie definite su uno stesso spazio di probabilità $(\Omega, \mathcal{F}, \mu)$, indicizzata da un parametro t che in genere rappresenta il tempo. Per ogni istante t , la variabile aleatoria $X_t : \Omega \rightarrow S$ assume valori in uno spazio di stati S , ma la vera entità da studiare non è il singolo X_t , bensì la mappa che a ogni $\omega \in \Omega$ associa la traiettoria $t \mapsto X_t(\omega)$. In altre parole, a livello ontologico, un processo è una misura μ su uno spazio di funzioni $\omega : I \rightarrow S$; e le variabili X_t non sono che proiezioni, coordinate temporali di questa misura globale. Si può pensare a X_t come a una funzione di valutazione $X_t(\omega) = \omega(t)$, e a μ come a una misura su Ω che, tramite l'identificazione di ogni ω con la sua traiettoria, diventa una misura su uno spazio di funzioni. È questa trasposizione a livello di cammini che segna il vero ingresso nel dominio della probabilità continua.

Nella forma più elementare, quando il tempo è discreto e indicizzato da $t \in \mathbb{N}$, un processo stocastico è una sequenza $\{X_0, X_1, X_2, \dots\}$. Ogni traiettoria è una sequenza $(x_0, x_1, x_2, \dots)$ e la probabilità μ è una misura definita su un prodotto infinito, tipicamente dotato della σ -algebra cilindrica generata dagli eventi che fissano un numero finito di coordinate. Già in questo contesto compaiono fenomeni profondi: leggi dei grandi numeri, processi di Markov, catene con distribuzioni stazionarie, convergenza verso equilibri. Ma il tempo discreto, per quanto ricco, mantiene un carattere “saltato”, granulare, in cui ogni aggiornamento è un passo netto. Il Volume III, al contrario, si concentra progressivamente su contesti in cui il tempo è un intervallo continuo, spesso $[0, \infty)$, e in cui la struttura della probabilità è intimamente intrecciata con l'analisi.

Quando il tempo è continuo, un processo stocastico $\{X_t : t \geq 0\}$ si può pensare come una variabile aleatoria a valori in uno spazio di funzioni $I \rightarrow S$, spesso dotato di una struttura topologica naturale (per esempio la topologia della convergenza uniforme sui compatti). Lo spazio dei cammini, come $C([0, T], \mathbb{R}^n)$, non è un semplice prodotto cartesiano, ma un oggetto dotato di una struttura topologica e metrica raffinata. Una misura su questo spazio, come la misura di Wiener che governa il moto browniano, è una misura che vive su un oggetto infinito-dimensionale, e la logica probabilistica deve imparare a parlare non di punti, ma di funzioni. Dire che “quasi sicuramente X_t è continuo in t ” significa che μ assegna misura 1 all'insieme dei cammini continui; dire che “quasi sicuramente X_t non è derivabile in nessun punto” significa che μ concentra tutta la sua massa su traiettorie che sono

ovunque irregolari, nonostante la continuità. Questa coesistenza di continuità e irregolarità è una delle firme del dominio che stiamo per esplorare.

Una delle prime strutture fondamentali da introdurre, non appena si considerano processi stocastici nel tempo, è la nozione di filtrazione. Una filtrazione $\{\mathcal{F}_t : t \geq 0\}$ è una famiglia crescente di σ -algebre, $\mathcal{F}_s \subseteq \mathcal{F}_t$ per $s \leq t$, che rappresenta l'informazione disponibile fino al tempo t . L'idea è che non osserviamo il processo nel suo complesso, ma lo vediamo svolgersi nel tempo, rivelando progressivamente nuove informazioni. Una variabile aleatoria X_t che descrive lo stato al tempo t deve essere misurabile rispetto a $\mathcal{F}_t$: questa condizione, chiamata adattamento, esprime formalmente che il valore di X_t è determinato dall'informazione disponibile fino a quel momento e non dipende da "anticipazioni" del futuro. L'introduzione di filtrazioni sposta la probabilità da un punto di vista statico a uno intrinsecamente dinamico: la probabilità non è solo una misura su Ω , ma una famiglia di visioni parziali, $\mathcal{F}_t$, che si affinano col tempo.

Con la filtrazione entra in scena una delle nozioni centrali della probabilità continua: la martingala. Un processo $\{M_t\}$ adattato a $\{\mathcal{F}_t\}$ è una martingala se, per ogni $s \leq t$, vale $E[M_t | \mathcal{F}_s] = M_s$ quasi sicuramente. Questa condizione esprime una forma forte di equità epistemica: alla luce di ciò che sappiamo al tempo s , il valore atteso di M_t , con $t \geq s$, è esattamente il valore corrente M_s . In altre parole, M_t non ha deriva prevedibile: nessuna strategia basata sull'informazione passata può produrre un guadagno atteso. Le martingale giocano, nella probabilità continua, un ruolo simile a quello delle ennuple ortogonali in algebra lineare o dei vettori tangenti in geometria: sono le strutture naturali sulle quali si costruisce la teoria. Molti processi fondamentali, dal moto browniano alle soluzioni di certe equazioni differenziali stocastiche, sono martingale o somma di martingale e processi di variazione finita.

Una proprietà profonda delle martingale è il teorema di arresto, o optional stopping. Se $\{M_t\}$ è una martingala e τ è un tempo di arresto, ossia una variabile aleatoria con valori in $[0, \infty]$ tale che $\{\tau \leq t\} \in \mathcal{F}_t$ per ogni t , allora sotto ipotesi tecniche adeguate si ha $E[M_\tau] = E[M_0]$. Questa affermazione, apparentemente semplice, esprime un principio di conservazione dell'aspettativa anche quando si sceglie il tempo di osservazione in modo dipendente dalla traiettoria. È come dire che, in un gioco equo, anche se decidi di fermarti quando "ti sembra conveniente", non puoi alterare l'equità attesa del gioco. L'optional stopping è una delle idee chiave che legano la probabilità continua alla logica del tempo e al concetto di strategia.

Per comprendere davvero cosa significhi "misura sui cammini", è utile focalizzarsi su un processo paradigmatico: il moto browniano. Il moto browniano $\{B_t : t \geq 0\}$ è definito (sotto una scelta di precisione tecnica) da tre proprietà fondamentali: $B_0 = 0$ quasi sicuramente; gli incrementi $B_t - B_s$ sono indipendenti per intervalli disgiunti e distribuiti come una normale $N(0, t-s)$; la funzione $t \mapsto B_t$ è continua quasi sicuramente. Queste tre condizioni determinano una misura μ_W , la misura di Wiener, sullo spazio $C([0, \infty), \mathbb{R})$. Si può pensare a μ_W come alla distribuzione "uniforme" su tutte le traiettorie continue, ma con una densità che privilegia i cammini che oscillano con una certa struttura gaussiana. La ricchezza di μ_W sta nel fatto che quasi ogni traiettoria è continua ma ha variazione quadratica crescente come t , mentre la variazione totale è infinita su ogni intervallo. Questo implica, tra le altre cose, che quasi ogni cammino non è derivabile in nessun punto: in ogni intervallo per quanto piccolo, il cammino oscilla in modo selvaggio.

Il moto browniano è un oggetto allo stesso tempo semplice e profondissimo. Semplice perché i suoi incrementi sono gaussiani indipendenti, il che rende molti calcoli relativamente accessibili. Profondissimo perché la sua struttura topologica e geometrica è estremamente ricca: il moto browniano esplora lo spazio in modo frattale, ha dimensione di Hausdorff pari a 2 in $\mathbb{R}^1$, ha tempi di

hitting distribuiti in modo complesso, ha massimi e minimi locali densi. Inoltre, il moto browniano è il limite di molte strutture discrete: passeggiate aleatorie normalizzate, modelli di diffusione, somme parziali di variabili indipendenti e identicamente distribuite. È il prototipo del rumore gaussiano continuo e il mattone fondamentale per costruire le equazioni differenziali stocastiche.

Una volta accettata l'idea che la probabilità viva su spazi di traiettorie, diventa naturale considerare trasformazioni di questi cammini e dinamiche indotte da equazioni che mescolano parti deterministiche e rumore. Le equazioni differenziali stocastiche, che saranno al centro dei capitoli successivi, nascono esattamente così: si considera un processo X_t che soddisfa un'equazione del tipo $dX_t = b(X_t, t) dt + \sigma(X_t, t) dB_t$, dove B_t è un moto browniano, b è un termine di deriva e σ è un coefficiente di diffusione. Questa equazione non va intesa nel senso classico, perché B_t non è derivabile; va intesa come un'equazione integrale: $X_t = X_0 + \int_0^t b(X_s, s) ds + \int_0^t \sigma(X_s, s) dB_s$, dove il secondo integrale è un integrale stocastico nel senso di Itô. Ma già a questo livello concettuale appare l'idea fondamentale: X_t è un funzionale del cammino B_t , e la sua distribuzione è la spinta in avanti della misura di Wiener tramite la mappa che associa a ogni cammino di B la corrispondente soluzione della SDE. La misura sui cammini di X non è altro che l'immagine di μ_W tramite un operatore deterministico (quando la SDE è ben posta e l'unicità forte è garantita).

Questa visione sottolinea un punto cruciale: nella probabilità continua, le distribuzioni sui cammini sono spesso ottenute come immagini (pushforward) di misure fondamentali come la misura di Wiener tramite trasformazioni deterministiche (soluzioni di SDE) o quasi-deterministiche (con elementi di controllo). La logica della probabilità non è più “assegno valori a eventi”, ma “definisco trasformazioni fra misure su spazi di funzioni”. Il linguaggio naturale per questo tipo di ragionamento non è più solo quello delle σ -algebre e delle misure, ma anche quello dei kernel stocastici, dei funtori probabilistici, delle coalgebre, che si intrecciano con la struttura analitica. Tuttavia, nel Volume III il baricentro si sposta: mentre nel Volume II queste trasformazioni erano viste quasi sempre in un quadro discreto o strutturale, qui sono immerse nella continuità, nella topologia e nell'analisi.

Un ulteriore aspetto delicatissimo, che emergerà più chiaramente man mano che si procede, è che la probabilità sui cammini forza a ripensare anche la nozione di evento come oggetto logico. In un modello classico, un evento è un sottoinsieme di stati; in un modello dinamico continuo, un evento può essere, ad esempio, “il cammino entra in un certo insieme A prima del tempo T ”, cioè l'evento $\{\omega : \exists t \leq T, X_t(\omega) \in A\}$. Questo evento non si riferisce a un valore puntuale, ma a un'intera storia. Il suo carattere temporale rende naturale considerare tempi di arresto, come $\tau_A = \inf\{t \geq 0 : X_t \in A\}$, che sono a loro volta variabili aleatorie che dipendono dal cammino intero fino al primo ingresso in A . Gli enunciati probabilistici diventano inevitabilmente enunciati sulla forma delle traiettorie, su tempi di hitting, su eventi di ritorno, su occupazioni nel tempo. La logica della probabilità sui cammini è una logica temporale e topologica prima ancora che proposizionale.

Infine, c'è un punto strutturale che merita di essere reso esplicito sin dall'inizio di questo volume: la probabilità sui cammini non è un semplice “caso speciale continuo” di quanto visto nel discreto, ma un cambio di regime concettuale. Nel discreto, si può sempre pensare a μ come a una media su mondi possibili indicizzati da n ; nel continuo, gli oggetti tipici sono cammini che, presi uno per uno, hanno probabilità zero, e ciò nonostante la loro struttura collettiva dà luogo a fenomeni stabili e misurabili. La misura di Wiener, per esempio, assegna probabilità zero a ogni singolo cammino, ma produce una teoria ricchissima di hitting times, massimi, tempi di ritorno, valori locali. Questa tensione permanente tra il fatto che “ogni cammino ha probabilità zero” e che “gli insiemi di cammini che realizzano certi comportamenti hanno probabilità positiva ben definita” è una delle caratteristiche distintive della probabilità continua e del pensiero analitico che la sostiene.

Il moto browniano e l'integrale di Itô

Il moto browniano rappresenta il punto di svolta concettuale in cui la probabilità cessa definitivamente di essere una teoria su insiemi di stati e diventa una teoria su spazi di funzioni. La sua natura è duplice: da un lato è il limite di processi discreti, come le passeggiate aleatorie normalizzate; dall'altro è un oggetto continuo che vive in uno spazio di cammini $C([0, \infty), \mathbb{R})$, dotato di una misura μ_W che non può essere compresa attraverso la sola intuizione combinatoria. Il moto browniano è la prima manifestazione di un rumore infinitesimale che non è derivabile, non è regolare, non è approssimabile da funzioni lisce, e tuttavia è governato da una struttura matematica estremamente rigida. È un oggetto che sfugge alla geometria classica, ma che obbedisce a una geometria probabilistica profonda, in cui la variazione quadratica sostituisce la derivata e in cui la misura di Wiener sostituisce la nozione di densità.

La definizione del moto browniano $\{B_t : t \geq 0\}$ si basa su tre proprietà fondamentali: $B_0=0$ quasi sicuramente; gli incrementi $B_t - B_s$ sono indipendenti e distribuiti come una variabile normale $N(0, t-s)$; la funzione $t \mapsto B_t$ è continua quasi sicuramente. Queste tre condizioni determinano univocamente la legge del processo, e la misura μ_W che ne deriva è una misura su $C([0, \infty), \mathbb{R})$ che assegna probabilità zero a ogni singolo cammino, ma che concentra la sua massa su un insieme di traiettorie che sono continue, irregolari, frattali, non derivabili in nessun punto. La continuità quasi sicura è una proprietà sorprendente: nonostante la totale irregolarità locale, il moto browniano non presenta salti. La sua irregolarità è infinitesimale, non macroscopica. Ogni cammino è un oggetto che oscilla in modo selvaggio su ogni intervallo, ma senza mai spezzarsi.

La struttura infinitesimale del moto browniano è catturata dalla nozione di variazione quadratica. Per una funzione liscia f , la variazione quadratica su un intervallo $[0, t]$ è zero; per il moto browniano, la variazione quadratica è t . Questo significa che, se si considera una partizione π_n dell'intervallo $[0, t]$ con passo massimo che tende a zero, la somma $\sum (B_{t_{i+1}} - B_{t_i})^2$ converge quasi sicuramente a t . Questa proprietà è la chiave per comprendere l'integrale di Itô. La variazione quadratica è ciò che sopravvive nel limite quando si moltiplicano incrementi infinitesimali del processo. È la firma analitica del rumore gaussiano continuo.

L'integrale di Itô nasce dalla necessità di definire un integrale rispetto a un processo che non è derivabile. Se si tenta di definire $\int_0^t \sigma_s dB_s$ come limite di somme $\sum \sigma_{t_i} (B_{t_{i+1}} - B_{t_i})$, si scopre che il limite esiste in L^2 e non dipende dalla scelta della partizione, purché si usino i valori σ_{t_i} presi all'inizio degli intervalli. Questa scelta non è arbitraria: è ciò che distingue l'integrale di Itô dall'integrale di Stratonovich, che usa valori medi e che obbedisce alla regola della catena classica. L'integrale di Itô, invece, obbedisce a una regola della catena modificata, la formula di Itô, che incorpora la variazione quadratica del moto browniano.

La formula di Itô è il cuore analitico della probabilità continua. Se X_t è un processo che soddisfa una SDE del tipo $dX_t = b(X_t, t) dt + \sigma(X_t, t) dB_t$, e f è una funzione sufficientemente regolare, allora $df(X_t) = f'(X_t) b(X_t, t) dt + f'(X_t) \sigma(X_t, t) dB_t + (1/2) f''(X_t) \sigma(X_t, t)^2 dt$. Il termine $(1/2) f'' \sigma^2 dt$ è la manifestazione della variazione quadratica del moto browniano. È ciò che distingue la dinamica stocastica dalla dinamica deterministica. È ciò che rende il rumore un oggetto infinitesimale ma non trascurabile. È ciò che permette alla probabilità di entrare nell'analisi in modo strutturale, non come perturbazione.

La formula di Itô permette di derivare l'equazione di Fokker-Planck, che descrive l'evoluzione della densità di X_t . Se $p(x, t)$ è la densità di X_t , allora $\partial_t p = -\partial_x(b p) + (1/2) \partial_{xx}(\sigma^2 p)$. Questa equazione parabolica è la controparte densitaria della SDE. La dinamica stocastica e la dinamica densitaria sono due facce della stessa struttura: una descrive l'evoluzione delle traiettorie, l'altra l'evoluzione delle distribuzioni. La probabilità continua è quindi una teoria unificata della dinamica dei cammini e della dinamica delle densità.

Il moto browniano è anche un oggetto geometrico. La sua traiettoria ha dimensione di Hausdorff pari a 2, nonostante viva in $\mathbb{R}^1$. Questo significa che il cammino riempie lo spazio in modo molto più denso di una curva liscia. La sua struttura frattale è evidente in ogni intervallo, per quanto piccolo. Il moto browniano è anche ricorrente in dimensione 1 e 2: ritorna arbitrariamente vicino al punto di partenza infinite volte. In dimensione 3 e superiore, invece, è transiente: si allontana e non ritorna più. Questa distinzione è una delle manifestazioni più profonde della geometria probabilistica.

Il moto browniano è un oggetto universale. È il limite di passeggiate aleatorie normalizzate, come afferma il teorema di Donsker. È il rumore fondamentale che appare in ogni limite centrale funzionale. È la componente gaussiana di ogni processo di Lévy. È il rumore che governa la diffusione termica, la fluttuazione quantistica nel limite classico, la dinamica di molte equazioni differenziali stocastiche. È il mattone fondamentale della probabilità continua.

L'integrale di Itô permette di definire le equazioni differenziali stocastiche, che saranno il tema del prossimo capitolo. Una SDE è un'equazione del tipo $dX_t = b(X_t, t) dt + \sigma(X_t, t) dB_t$, e la sua soluzione è un processo X_t che è un funzionale del cammino B_t . La misura sui cammini di X è l'immagine della misura di Wiener tramite la mappa che associa a ogni cammino di B la soluzione della SDE. La probabilità continua è quindi una teoria delle trasformazioni di misure su spazi di funzioni.

Il moto browniano e l'integrale di Itô costituiscono il nucleo analitico del Volume III. Essi rappresentano la prima apparizione del rumore come oggetto infinitesimale ma strutturato, della variazione quadratica come sostituto della derivata, della misura di Wiener come misura fondamentale su spazi di cammini, della dinamica stocastica come trasformazione di misure. Tutto ciò che seguirà — le SDE, le equazioni di Fokker–Planck, la teoria delle martingale, la teoria ergodica stocastica, la geometria probabilistica — si innesta su questa base.

Equazioni differenziali stocastiche

Le equazioni differenziali stocastiche rappresentano il punto in cui la probabilità continua si fonde con la dinamica, dando origine a una teoria in cui il caso non è più un disturbo esterno, ma una componente strutturale dell'evoluzione. Una SDE, nella sua forma più elementare, è un'equazione del tipo $dX_t = b(X_t, t) dt + \sigma(X_t, t) dB_t$, dove B_t è un moto browniano, b è un termine di deriva e σ è un coefficiente di diffusione. Questa scrittura, apparentemente simile a quella delle equazioni differenziali ordinarie, nasconde una differenza concettuale radicale: il termine dB_t non è un differenziale nel senso classico, perché B_t non è derivabile. È un oggetto infinitesimale che possiede variazione quadratica, e la sua presenza modifica profondamente la natura dell'equazione. La SDE non è un'equazione differenziale nel senso classico, ma un'equazione integrale in cui l'integrale rispetto a B_t è un integrale di Itô.

Per comprendere la natura di una SDE, è utile partire dalla sua interpretazione integrale. Una soluzione X_t è un processo adattato alla filtrazione generata da B_t che soddisfa $X_t = X_0 + \int_0^t b(X_s, s) ds + \int_0^t \sigma(X_s, s) dB_s$. La prima parte, $\int b ds$, è un integrale classico, che rappresenta la componente deterministica dell'evoluzione. La seconda parte, $\int \sigma dB$, è un integrale stocastico, che rappresenta la componente aleatoria. La presenza di questo secondo termine rende X_t un funzionale del cammino B_t , e la sua distribuzione è l'immagine della misura di Wiener tramite la mappa che associa a ogni cammino di B la soluzione della SDE. La SDE è quindi una trasformazione di misure su spazi di cammini.

La teoria dell'esistenza e unicità delle soluzioni di una SDE si basa su condizioni di Lipschitz e di crescita lineare sui coefficienti b e σ . Se b e σ sono Lipschitz in x e misurabili in t , allora esiste

un'unica soluzione forte, cioè un processo X_t definito sullo stesso spazio di probabilità del moto browniano e adattato alla sua filtrazione. Questa unicità forte è fondamentale: significa che la soluzione è determinata dal cammino del moto browniano e dalla condizione iniziale. La SDE è quindi una trasformazione deterministica del rumore. Se invece si considera l'unicità debole, si richiede solo che la legge del processo sia unica, non necessariamente la soluzione stessa. Questa distinzione tra unicità forte e debole è una delle caratteristiche più sottili della teoria delle SDE.

La presenza del termine $\sigma(X_t, t) dB_t$ introduce una componente di diffusione che rende la dinamica intrinsecamente irregolare. Anche se b e σ sono funzioni lisce, la soluzione X_t non è liscia: è continua ma non derivabile, e la sua irregolarità è governata dalla variazione quadratica del moto browniano. La SDE è quindi una dinamica che vive su uno spazio di funzioni irregolari. La regolarità di X_t dipende dalla regolarità di σ , ma non può superare quella del moto browniano. La SDE è una dinamica che si sviluppa su un substrato frattale.

La formula di Itô permette di analizzare la dinamica di funzioni di X_t . Se f è una funzione sufficientemente regolare, allora $df(X_t) = f'(X_t) b(X_t, t) dt + f'(X_t) \sigma(X_t, t) dB_t + (1/2) f''(X_t) \sigma(X_t, t)^2 dt$. Questo risultato è la chiave per comprendere la relazione tra la SDE e l'equazione di Fokker-Planck. Se $p(x, t)$ è la densità di X_t , allora p soddisfa l'equazione $\partial_t p = -\partial_x(b p) + (1/2) \partial_{xx}(\sigma^2 p)$. Questa equazione descrive l'evoluzione della densità nel tempo. La SDE e l'equazione di Fokker-Planck sono due rappresentazioni della stessa dinamica: una descrive l'evoluzione delle traiettorie, l'altra l'evoluzione delle distribuzioni. La probabilità continua è quindi una teoria unificata della dinamica dei cammini e della dinamica delle densità.

La SDE è una teoria della semimartingala. Una semimartingala è un processo che può essere scritto come somma di un processo di variazione finita e di una martingala locale. Le soluzioni delle SDE sono semimartingale, e la teoria delle semimartingale fornisce il contesto naturale per definire l'integrale di Itô. La SDE è quindi una dinamica che vive nello spazio delle semimartingale, che è lo spazio naturale della probabilità continua.

La SDE è una teoria della trasformazione del rumore. Il moto browniano è il rumore fondamentale, e la SDE è una trasformazione di questo rumore tramite i coefficienti b e σ . La misura sui cammini di X è l'immagine della misura di Wiener tramite questa trasformazione. La SDE è quindi una teoria della misura su spazi di funzioni, in cui la dinamica è una trasformazione di misure.

La SDE è anche una teoria della dipendenza dal rumore. Se si considera una SDE con coefficiente σ identicamente nullo, si ottiene un'equazione differenziale ordinaria. Se σ è non nullo, la dinamica è influenzata dal rumore. La presenza del rumore può rendere la dinamica più regolare o più irregolare, a seconda della struttura di b e σ . In alcuni casi, il rumore può stabilizzare la dinamica, come nel fenomeno della stabilizzazione stocastica. In altri casi, può destabilizzarla. La SDE è quindi una teoria della stabilità e dell'instabilità in presenza di rumore.

La SDE è, in più, una teoria della dipendenza dalla condizione iniziale. Se si considera una famiglia di soluzioni X_t^x con condizione iniziale $X_0 = x$, si ottiene un flusso stocastico. Questo flusso non è un flusso deterministico, perché le traiettorie non sono invertibili. Tuttavia, esiste una teoria dei flussi stocastici che permette di descrivere la dipendenza della soluzione dalla condizione iniziale. La SDE è quindi una teoria della geometria stocastica.

Infine, la SDE è una teoria della misura invariata. Una misura π è invariante per la SDE se la distribuzione di X_t è π per ogni t . Questa condizione è equivalente al fatto che π sia una soluzione stazionaria dell'equazione di Fokker-Planck. La SDE è quindi una teoria dell'equilibrio probabilistico.

Le equazioni differenziali stocastiche rappresentano quindi una generalizzazione profonda delle equazioni differenziali ordinarie. Esse incorporano il rumore come componente strutturale, trasformano la dinamica in una trasformazione di misure su spazi di cammini, unificano la dinamica delle traiettorie e la dinamica delle densità, vivono nello spazio delle semimartingale, descrivono la stabilità e l'instabilità in presenza di rumore, definiscono flussi stocastici e misure invarianti. La SDE è il cuore della probabilità continua, e tutto ciò che seguirà — teoria ergodica stocastica, teoria dei filtri, teoria delle misure invarianti, geometria stocastica — si innesta su questa struttura.

Equazioni di Fokker–Planck e dinamica delle densità

L'equazione di Fokker–Planck rappresenta il punto in cui la dinamica stocastica, descritta a livello di traiettorie dalle equazioni differenziali stocastiche, si traduce in una dinamica delle densità, cioè in un'evoluzione deterministica della distribuzione di probabilità. È il luogo in cui la probabilità continua rivela la sua duplice natura: da un lato teoria dei cammini, dall'altro teoria delle misure. La SDE $dX_t = b(X_t, t) dt + \sigma(X_t, t) dB_t$ descrive l'evoluzione di un singolo cammino; l'equazione di Fokker–Planck $\partial_t p = -\partial_x(b p) + (1/2) \partial_{xx}(\sigma^2 p)$ descrive l'evoluzione della densità $p(x, t)$ della distribuzione di X_t . Le due descrizioni sono equivalenti, ma vivono in spazi concettuali diversi: la prima nello spazio dei cammini, la seconda nello spazio delle misure. La probabilità continua è la teoria che unifica queste due prospettive.

Per comprendere l'origine dell'equazione di Fokker–Planck, è utile partire dalla formula di Itô. Se X_t è una soluzione della SDE e f è una funzione sufficientemente regolare, allora $df(X_t) = f'(X_t) b(X_t, t) dt + f'(X_t) \sigma(X_t, t) dB_t + (1/2) f''(X_t) \sigma(X_t, t)^2 dt$. Prendendo l'aspettativa, il termine stocastico scompare, perché l'integrale di Itô ha valore atteso zero. Si ottiene quindi $d/dt E[f(X_t)] = E[f'(X_t) b(X_t, t)] + (1/2) E[f''(X_t) \sigma(X_t, t)^2]$. Se X_t ha densità $p(x, t)$, questa identità può essere riscritta come $\int f(x) \partial_t p(x, t) dx = \int f(x) b(x, t) p(x, t) dx + (1/2) \int f''(x) \sigma(x, t)^2 p(x, t) dx$. Integrando per parti e usando la regolarità di f , si ottiene $\int f(x) \partial_t p(x, t) dx = \int f(x) [-\partial_x(b p) + (1/2) \partial_{xx}(\sigma^2 p)] dx$. Poiché questa identità vale per ogni f , si ottiene l'equazione di Fokker–Planck $\partial_t p = -\partial_x(b p) + (1/2) \partial_{xx}(\sigma^2 p)$. Questa equazione è la controparte densitaria della SDE. La dinamica stocastica e la dinamica densitaria sono due rappresentazioni della stessa evoluzione.

L'equazione di Fokker–Planck è un'equazione parabolica, simile all'equazione del calore. Il termine $(1/2) \partial_{xx}(\sigma^2 p)$ rappresenta la diffusione, cioè la tendenza della distribuzione a spandersi nel tempo. Il termine $-\partial_x(b p)$ rappresenta la deriva, cioè la tendenza della distribuzione a muoversi nella direzione determinata da b . La combinazione di questi due effetti produce una dinamica complessa, in cui la distribuzione può spostarsi, allargarsi, contrarsi, stabilizzarsi. La SDE descrive la dinamica di un singolo cammino; l'equazione di Fokker–Planck descrive la dinamica dell'intera distribuzione.

La relazione tra SDE e Fokker–Planck è profonda. La SDE è una dinamica su cammini; la Fokker–Planck è una dinamica su misure. La SDE è una trasformazione della misura di Wiener; la Fokker–Planck è una trasformazione della densità. La SDE è una dinamica non deterministica; la Fokker–Planck è una dinamica deterministica. La SDE è una teoria della variazione quadratica; la Fokker–Planck è una teoria della diffusione. La SDE è una teoria della semimartingala; la Fokker–Planck è una teoria delle equazioni paraboliche. La probabilità continua è la teoria che unifica queste due prospettive.

L'equazione di Fokker–Planck è anche una teoria dell'equilibrio. Una densità $\pi(x)$ è stazionaria se $\partial_t \pi = 0$, cioè se $-\partial_x(b \pi) + (1/2) \partial_{xx}(\sigma^2 \pi) = 0$. Questa equazione descrive le distribuzioni invarianti della SDE. Se la SDE è ergodica, la distribuzione di X_t converge a π per $t \rightarrow \infty$. La

Fokker–Planck è quindi una teoria dell’equilibrio probabilistico. La SDE descrive il cammino verso l’equilibrio; la Fokker–Planck descrive la convergenza della distribuzione.

L’equazione di Fokker–Planck è anche una teoria della reversibilità. Una distribuzione π è reversibile se soddisfa la condizione di bilancio dettagliato $b \pi = (1/2) \partial_x(\sigma^2 \pi)$. In questo caso, la SDE è reversibile nel senso che il processo ha la stessa legge se il tempo viene invertito. La reversibilità è una proprietà profonda, che collega la dinamica stocastica alla termodinamica. La Fokker–Planck è quindi una teoria della simmetria temporale.

L’equazione di Fokker–Planck è anche una teoria della dissipazione. La diffusione tende ad aumentare l’entropia della distribuzione. Se si considera l’entropia $H(t) = -\int p(x,t) \log p(x,t) dx$, si può mostrare che $dH/dt \geq 0$ sotto condizioni appropriate. La Fokker–Planck è quindi una teoria della dissipazione dell’informazione. La SDE descrive la fluttuazione; la Fokker–Planck descrive la dissipazione.

L’equazione di Fokker–Planck costituisce una teoria della geometria. La diffusione può essere interpretata come un flusso su una varietà, e la Fokker–Planck può essere generalizzata a varietà riemanniane sostituendo ∂_{xx} con il laplaciano di Laplace–Beltrami. La SDE diventa una SDE su varietà, e la Fokker–Planck diventa un’equazione parabolica su una varietà. La probabilità continua è quindi una teoria della geometria stocastica.

Infine, l’equazione di Fokker–Planck è una teoria della misura. La SDE descrive la trasformazione della misura di Wiener; la Fokker–Planck descrive la trasformazione della densità. La SDE è una teoria della misura su cammini; la Fokker–Planck è una teoria della misura su spazi di stati. La probabilità continua è la teoria che unifica queste due prospettive.

L’equazione di Fokker–Planck rappresenta quindi una delle strutture fondamentali della probabilità continua. Essa unifica la dinamica stocastica e la dinamica densitaria, descrive l’equilibrio, la reversibilità, la dissipazione, la geometria, la misura. È la controparte analitica delle equazioni differenziali stocastiche, e tutto ciò che seguirà — teoria ergodica stocastica, teoria dei filtri, teoria delle misure invarianti, geometria stocastica — si innesta su questa struttura.

Teoria ergodica stocastica

La teoria ergodica stocastica rappresenta il punto in cui la probabilità continua incontra la nozione di tempo infinito, di equilibrio, di ricorrenza e di stabilità. È la parte della teoria dei processi stocastici che studia ciò che accade quando il tempo tende a infinito, non in senso puntuale, ma in senso distribuzionale, funzionale, misurale. Una SDE o un processo di Markov non sono solo dinamiche locali, ma sistemi che evolvono nel tempo e che, sotto condizioni appropriate, convergono verso una distribuzione stazionaria π , che rappresenta l’equilibrio statistico del sistema. La teoria ergodica stocastica è la teoria di questa convergenza, della sua velocità, della sua stabilità, della sua unicità, della sua struttura geometrica.

Per comprendere la natura dell’ergodicità, è utile partire dalla nozione di misura invariante. Una misura π è invariante per un processo X_t se la distribuzione di X_t è π per ogni t , quando X_0 è distribuito secondo π . Nel caso di una SDE $dX_t = b(X_t) dt + \sigma(X_t) dB_t$, la misura π è invariante se soddisfa l’equazione stazionaria di Fokker–Planck, cioè $0 = -\partial_x(b \pi) + (1/2) \partial_{xx}(\sigma^2 \pi)$. Questa equazione descrive l’equilibrio tra deriva e diffusione. La misura π è un punto fisso della dinamica densitaria. La sua esistenza e unicità dipendono dalla struttura di b e σ . Se la deriva spinge il processo verso una regione compatta e la diffusione non è troppo forte, la misura invariante esiste ed è unica.

Se la deriva è troppo debole o la diffusione troppo forte, il processo può divergere e non avere equilibrio.

La misura invariante è il primo passo verso l'ergodicità. Un processo è ergodico se, per ogni funzione integrabile f , la media temporale $(1/T) \int_0^T f(X_t) dt$ converge quasi sicuramente alla media spaziale $\int f d\pi$. Questa proprietà è profonda: significa che il tempo e lo spazio si equivalgono. La media lungo una singola traiettoria coincide con la media rispetto alla misura invariante. L'ergodicità è una forma di legge dei grandi numeri nel tempo continuo. È la proprietà che permette di stimare la misura invariante osservando una singola traiettoria. È la proprietà che rende la misura invariante non solo un oggetto teorico, ma un oggetto empirico.

La teoria ergodica stocastica è anche una teoria della ricorrenza. Un processo è ricorrente se ritorna arbitrariamente vicino al punto di partenza infinite volte. Il moto browniano in dimensione 1 e 2 è ricorrente; in dimensione 3 e superiore è transiente. Questa distinzione è una delle manifestazioni più profonde della geometria probabilistica. La ricorrenza è legata alla struttura del laplaciano e alla dimensione dello spazio. La diffusione in dimensione bassa è troppo forte per permettere al processo di allontanarsi definitivamente; in dimensione alta, la diffusione è sufficiente per permettere al processo di divergere. La ricorrenza è una proprietà globale del processo, che non può essere dedotta dalla sola dinamica locale.

La teoria ergodica stocastica è una teoria della velocità di convergenza. Se un processo è ergodico, la distribuzione di X_t converge a π per $t \rightarrow \infty$. Ma la velocità di questa convergenza può essere lenta o veloce. Se la convergenza è esponenziale, si parla di ergodicità geometrica. Questa proprietà è legata alla presenza di una forza di richiamo nella deriva b . Se b spinge il processo verso una regione compatta con forza sufficiente, la convergenza è esponenziale. Se b è troppo debole, la convergenza può essere polinomiale o addirittura più lenta. La velocità di convergenza è una proprietà cruciale per le applicazioni, perché determina quanto rapidamente il processo dimentica la condizione iniziale.

La teoria ergodica stocastica è una teoria della mescolanza. Un processo è mescolante se la correlazione tra $f(X_t)$ e $g(X_0)$ tende a zero per $t \rightarrow \infty$. Questa proprietà è più forte dell'ergodicità. L'ergodicità richiede che le medie temporali convergano; la mescolanza richiede che il processo perda memoria del passato. La mescolanza è una forma di indipendenza asintotica. È una proprietà che permette di applicare teoremi limite, come il teorema del limite centrale, ai processi stocastici. La mescolanza è una proprietà che dipende dalla struttura della deriva e della diffusione. Se la deriva è forte e la diffusione moderata, la mescolanza è rapida; se la deriva è debole, la mescolanza può essere lenta.

La teoria ergodica stocastica è anche una teoria della decomposizione. Un processo può avere più misure invarianti, e in questo caso non è ergodico. La presenza di più misure invarianti indica che il processo ha più regioni stabili, tra le quali non può passare. La decomposizione ergodica afferma che ogni misura invariante può essere decomposta in misure ergodiche. Questa decomposizione è analoga alla decomposizione spettrale di un operatore. La teoria ergodica stocastica è quindi una teoria della struttura interna delle misure invarianti.

La teoria ergodica stocastica è, ancora, una teoria della geometria del potenziale. La misura invariante π può essere scritta come $\pi(x) \propto e^{-V(x)}$ se la SDE è reversibile. In questo caso, V è un potenziale e la SDE descrive un moto browniano in un paesaggio di energia. La teoria ergodica stocastica diventa una teoria della dinamica in un potenziale. La profondità dei pozzi determina la velocità di convergenza, la presenza di barriere determina la metastabilità, la struttura del potenziale determina la decomposizione ergodica. La probabilità continua diventa una teoria della dinamica in un paesaggio energetico.

La teoria ergodica stocastica è, in aggiunta, una teoria della metastabilità. Se il potenziale ha più minimi profondi separati da barriere alte, il processo può rimanere per tempi lunghissimi in una regione prima di passare a un'altra. La distribuzione sembra stazionaria su scale temporali moderate, ma non lo è su scale temporali molto lunghe. La metastabilità è una proprietà fondamentale dei sistemi complessi. La teoria ergodica stocastica fornisce gli strumenti per analizzare questi fenomeni.

Infine, la teoria ergodica stocastica è una teoria della misura su tempi infiniti. La SDE descrive la dinamica locale; la Fokker–Planck descrive la dinamica densitaria; la teoria ergodica descrive la dinamica asintotica. È la parte della probabilità continua che studia ciò che accade quando il tempo tende a infinito. È la teoria dell'equilibrio, della ricorrenza, della mescolanza, della metastabilità, della decomposizione, della velocità di convergenza. È la teoria che permette di comprendere la struttura globale dei processi stocastici.

La teoria ergodica stocastica rappresenta quindi uno dei vertici concettuali della probabilità continua. Essa unifica la dinamica locale e la dinamica globale, la misura e la densità, il cammino e l'equilibrio, la diffusione e la deriva, la geometria e la termodinamica. È la teoria che permette di comprendere il comportamento a lungo termine dei sistemi stocastici, e tutto ciò che seguirà — teoria dei filtri, teoria delle misure invarianti, geometria stocastica, teoria dei campi — si innesta su questa struttura.

Filtri stocastici e teoria dell'informazione continua

La teoria dei filtri stocastici rappresenta il punto in cui la probabilità continua incontra la nozione di informazione parziale, di osservazione rumorosa, di inferenza dinamica. È la parte della teoria dei processi stocastici che studia come un agente, dotato di una sequenza continua di osservazioni imperfette, possa ricostruire, stimare o prevedere lo stato di un sistema che evolve secondo una dinamica stocastica. Se la SDE descrive l'evoluzione reale del sistema, il filtro descrive l'evoluzione della conoscenza dell'agente. La probabilità continua diventa così una teoria dell'informazione nel tempo, una teoria della conoscenza che si aggiorna in modo infinitesimale, una teoria della credenza che evolve sotto l'effetto simultaneo della dinamica e dell'osservazione.

Per comprendere la natura del filtraggio, è utile partire da un modello elementare. Si consideri un processo nascosto X_t che evolve secondo una SDE $dX_t = b(X_t, t) dt + \sigma(X_t, t) dB_t$, e un processo osservato Y_t che evolve secondo $dY_t = h(X_t, t) dt + v dW_t$, dove W_t è un moto browniano indipendente da B_t . L'agente osserva Y_t , non X_t . L'obiettivo è determinare la distribuzione condizionata $\pi_t(dx) = \mu(X_t \in dx \mid \mathcal{F}_t^Y)$, dove $\mathcal{F}_t^Y$ è la filtrazione generata dalle osservazioni Y_s per $s \leq t$. Questa distribuzione π_t è la credenza dell'agente sullo stato X_t alla luce delle osservazioni. La sua evoluzione nel tempo è il filtro stocastico.

La difficoltà concettuale è che π_t non è una variabile aleatoria a valori in $\mathbb{R}$, ma una variabile aleatoria a valori nello spazio delle misure di probabilità. La sua evoluzione è una SDE nello spazio delle misure. La teoria dei filtri stocastici è quindi una teoria della dinamica delle misure condizionate. La probabilità continua diventa una teoria della misura su spazi di misure.

La forma più elementare del filtro è il filtro di Kalman, che si applica quando la dinamica è lineare e il rumore è gaussiano. In questo caso, X_t e Y_t sono gaussiani, e la distribuzione condizionata π_t è anch'essa gaussiana. La sua evoluzione è descritta da un sistema di equazioni differenziali per la media m_t e la covarianza P_t . La media soddisfa $dm_t = b(m_t, t) dt + K_t (dY_t - h(m_t, t) dt)$, dove K_t è il guadagno di Kalman, e la covarianza soddisfa $dP_t/dt = A_t P_t + P_t A_t^T + \Sigma_t - K_t R_t K_t^T$. Il filtro di Kalman è una teoria della conoscenza lineare: la credenza è una gaussiana che si aggiorna in modo ottimale alla luce delle osservazioni.

Ma la teoria generale del filtraggio è molto più profonda. Quando la dinamica non è lineare o il rumore non è gaussiano, la distribuzione condizionata π_t non è più gaussiana. La sua evoluzione è descritta dall'equazione di Kushner–Stratonovich, che è una SDE nello spazio delle misure: $d\pi_t(f) = \pi_t(\mathcal{L}f) dt + [\pi_t(fh) - \pi_t(f)\pi_t(h)] (dY_t - \pi_t(h) dt) / v^2$, dove $\mathcal{L}$ è il generatore infinitesimale della SDE di X_t . Questa equazione descrive l'evoluzione della credenza dell'agente. Il primo termine, $\pi_t(\mathcal{L}f) dt$, rappresenta l'effetto della dinamica; il secondo termine rappresenta l'effetto dell'osservazione. La credenza si aggiorna in modo continuo, infinitesimale, sotto l'effetto simultaneo della dinamica e dell'informazione.

La forma più elegante dell'equazione del filtro è l'equazione di Zakai, che descrive l'evoluzione della densità non normalizzata ρ_t : $d\rho_t = \mathcal{L}^* \rho_t dt + (h \rho_t / v^2) dY_t$. Questa equazione è lineare, a differenza dell'equazione di Kushner–Stratonovich, che è non lineare. La normalizzazione $\pi_t = \rho_t / \int \rho_t dx$ restituisce la distribuzione condizionata. L'equazione di Zakai è una SDE lineare nello spazio delle misure, e rappresenta la forma più analitica del filtraggio.

La teoria dei filtri stocastici è una teoria dell'informazione. Il processo osservato Y_t contiene informazione su X_t , ma questa informazione è distorta dal rumore. Il filtro è il meccanismo che estrae l'informazione utile. La quantità $dY_t - \pi_t(h) dt$ è l'innovazione, cioè la parte dell'osservazione che non è prevista dal modello. L'innovazione è un moto browniano rispetto alla filtrazione π_t . Questa proprietà è profonda: significa che il filtro trasforma l'osservazione rumorosa in un rumore bianco “puro”, che rappresenta l'informazione nuova. La teoria dei filtri stocastici è quindi una teoria della decomposizione dell'informazione in parte prevista e parte innovativa.

La teoria dei filtri stocastici è una teoria della stabilità. Se due agenti osservano la stessa sequenza Y_t ma partono da credenze iniziali diverse, le loro credenze π_t tendono a convergere. Questa proprietà è chiamata stabilità del filtro. Essa significa che l'informazione contenuta nelle osservazioni è sufficiente per cancellare l'effetto della credenza iniziale. La stabilità è una proprietà cruciale per l'identificazione dei sistemi. Se il filtro è stabile, l'agente può apprendere la dinamica del sistema osservando Y_t .

La teoria dei filtri stocastici è una teoria della non linearità. Quando la dinamica è non lineare, la distribuzione condizionata π_t può essere multimodale, asimmetrica, complessa. La sua evoluzione è una SDE nello spazio delle misure, e la sua analisi richiede strumenti avanzati di analisi funzionale. La teoria dei filtri stocastici è quindi una teoria della dinamica delle misure in spazi infiniti.

La teoria dei filtri stocastici è anche una teoria della geometria dell'informazione. La distribuzione condizionata π_t può essere vista come un punto in uno spazio di probabilità dotato di una metrica naturale, come la metrica di Fisher o la metrica di Wasserstein. L'evoluzione di π_t è una traiettoria in questo spazio. La teoria dei filtri stocastici è quindi una teoria della geometria dell'informazione nel tempo.

Infine, la teoria dei filtri stocastici è dopotutto una teoria della conoscenza. La SDE descrive la realtà; il filtro descrive la conoscenza dell'agente. La SDE è una dinamica oggettiva; il filtro è una dinamica soggettiva. La SDE è una teoria del mondo; il filtro è una teoria della percezione. La probabilità continua diventa una teoria della conoscenza che si aggiorna in modo infinitesimale, una teoria dell'informazione che si accumula nel tempo, una teoria della credenza che evolve sotto l'effetto simultaneo della dinamica e dell'osservazione.

La teoria dei filtri stocastici rappresenta quindi uno dei vertici concettuali della probabilità continua. Essa unifica la dinamica stocastica e la dinamica dell'informazione, la misura e la credenza, la realtà e la percezione, la diffusione e l'osservazione, la geometria e l'inferenza. È la teoria che permette di

comprendere come un agente possa conoscere un mondo che evolve in modo stocastico, e tutto ciò che seguirà — teoria delle misure invarianti, geometria stocastica, teoria dei campi — si innesta su questa struttura.

Misure invarianti, potenziali stocastici e paesaggi energetici

La nozione di misura invariante rappresenta uno dei punti più profondi della probabilità continua, perché è il luogo in cui la dinamica infinitesimale delle equazioni differenziali stocastiche incontra la struttura globale dello spazio degli stati. Una misura π è invariante per un processo stocastico X_t se, iniziando il processo con distribuzione π , la distribuzione rimane π per ogni t . Questa definizione, apparentemente semplice, nasconde una struttura concettuale estremamente ricca: π è un punto fisso della dinamica densitaria, una soluzione stazionaria dell'equazione di Fokker–Planck, un equilibrio termodinamico, un attrattore misurale, una forma di simmetria temporale, un oggetto geometrico che codifica la struttura profonda della deriva e della diffusione. La misura invariante è il luogo in cui la dinamica locale si cristallizza in una struttura globale.

Per comprendere la natura di π , è utile partire dalla SDE $dX_t = b(X_t) dt + \sigma(X_t) dB_t$. La densità $p(x,t)$ della distribuzione di X_t soddisfa l'equazione di Fokker–Planck $\partial_t p = -\partial_x(b p) + (1/2) \partial_{xx}(\sigma^2 p)$. Una misura π è invariante se la sua densità $\pi(x)$ soddisfa l'equazione stazionaria $0 = -\partial_x(b \pi) + (1/2) \partial_{xx}(\sigma^2 \pi)$. Questa equazione è un bilancio tra deriva e diffusione. La deriva tende a spingere la massa verso certe regioni; la diffusione tende a spandere la massa. L'equilibrio tra questi due effetti determina la forma di π . La misura invariante è quindi una soluzione di un'equazione differenziale che rappresenta l'equilibrio tra forze deterministiche e fluttuazioni stocastiche.

Quando la SDE è reversibile, cioè quando esiste un potenziale V tale che $b = -(1/2) \sigma^2 \partial_x V$, la misura invariante ha la forma $\pi(x) \propto e^{-V(x)}$. Questa forma è la stessa della distribuzione di Gibbs nella meccanica statistica. Il potenziale V rappresenta un paesaggio energetico, e la misura invariante è la distribuzione di equilibrio di un sistema che si muove in questo paesaggio sotto l'effetto del rumore. La SDE descrive un moto browniano con deriva determinata dal gradiente del potenziale. La misura invariante è la distribuzione di equilibrio termico. La probabilità continua diventa una teoria della termodinamica stocastica.

Il paesaggio energetico V determina la struttura globale della dinamica. I minimi di V sono regioni stabili; i massimi sono regioni instabili; le barriere tra i minimi determinano la difficoltà del processo nel passare da una regione all'altra. Se V ha un unico minimo globale, la misura invariante è unica e il processo è ergodico. Se V ha più minimi separati da barriere alte, il processo può essere metastabile: può rimanere per tempi lunghissimi in una regione prima di passare a un'altra. La metastabilità è una proprietà fondamentale dei sistemi complessi. La misura invariante esiste ed è unica, ma il processo può impiegare tempi astronomici per esplorare tutte le regioni. La teoria ergodica stocastica descrive la convergenza asintotica; la teoria della metastabilità descrive la dinamica su scale temporali intermedie.

La misura invariante è una teoria della simmetria temporale. Se il processo è reversibile, la misura invariante soddisfa la condizione di bilancio dettagliato $b \pi = (1/2) \partial_x(\sigma^2 \pi)$. Questa condizione implica che il processo ha la stessa legge se il tempo viene invertito. La reversibilità è una proprietà profonda, che collega la dinamica stocastica alla termodinamica. Se il processo non è reversibile, la misura invariante può comunque esistere, ma la dinamica non è simmetrica nel tempo. La misura invariante è quindi una teoria della simmetria e della sua rottura.

La misura invariante è una teoria della geometria. Se la SDE è definita su una varietà riemanniana, la misura invariante è una misura su questa varietà, e la sua densità è determinata dal laplaciano di

Laplace–Beltrami e dal potenziale. La diffusione è governata dalla geometria della varietà; la deriva è governata dal potenziale. La misura invariante è quindi una teoria della geometria stocastica. La probabilità continua diventa una teoria della misura su varietà.

La misura invariante è una teoria della decomposizione. Se il processo ha più misure invarianti, esso non è ergodico. La presenza di più misure invarianti indica che il processo ha più regioni stabili tra le quali non può passare. La decomposizione ergodica afferma che ogni misura invariante può essere decomposta in misure ergodiche. Questa decomposizione è analoga alla decomposizione spettrale di un operatore. La misura invariante è quindi una teoria della struttura interna della dinamica.

La misura invariante è anche una teoria della stabilità. Se la SDE è ergodica, la distribuzione di X_t converge a π per $t \rightarrow \infty$. La velocità di questa convergenza dipende dalla struttura del potenziale e della diffusione. Se il potenziale ha un unico minimo profondo, la convergenza è rapida; se il potenziale ha più minimi separati da barriere alte, la convergenza può essere lenta. La misura invariante è quindi una teoria della stabilità e dell'instabilità.

Infine, la misura invariante è una teoria della misura su tempi infiniti. La SDE descrive la dinamica locale; la Fokker–Planck descrive la dinamica densitaria; la teoria ergodica descrive la dinamica asintotica; la misura invariante descrive l'equilibrio. La probabilità continua è la teoria che unifica queste quattro prospettive. La misura invariante è il punto in cui la dinamica locale si cristallizza in una struttura globale, in cui il rumore infinitesimale produce un equilibrio macroscopico, in cui la geometria del potenziale determina la struttura della distribuzione.

La misura invariante, il potenziale stocastico e il paesaggio energetico rappresentano quindi uno dei vertici concettuali della probabilità continua. Essi unificano la dinamica stocastica, la termodinamica, la geometria, la teoria delle misure, la teoria ergodica, la metastabilità. Sono la chiave per comprendere la struttura globale dei sistemi stocastici, e tutto ciò che seguirà — teoria dei campi stocastici, teoria delle misure su spazi funzionali, geometria stocastica avanzata — si innesta su questa struttura.

Geometria stocastica e SDE su varietà riemanniane

La geometria stocastica rappresenta il punto in cui la probabilità continua incontra la struttura geometrica dello spazio in cui i processi vivono. Finora abbiamo considerato processi stocastici definiti su $\mathbb{R}^n$, uno spazio lineare, dotato di una metrica euclidea e di una struttura differenziale semplice. Ma molti sistemi fisici, biologici, economici, informativi, non vivono in uno spazio lineare: vivono su varietà, su superfici curve, su spazi con vincoli, su gruppi di Lie, su orbifolds, su spazi di configurazioni. La dinamica stocastica deve allora adattarsi alla geometria dello spazio, e il rumore deve essere trasportato lungo la struttura differenziale della varietà. La probabilità continua diventa una teoria della diffusione su spazi curvi, una teoria della misura su varietà, una teoria della dinamica stocastica immersa nella geometria.

Per comprendere la natura di una SDE su una varietà riemanniana (M, g) , è utile partire dal moto browniano. In $\mathbb{R}^n$, il moto browniano è definito come il processo con incrementi gaussiani indipendenti e varianza proporzionale al tempo. Su una varietà, questa definizione non ha senso, perché non esiste una struttura lineare globale. La definizione corretta è che il moto browniano su (M, g) è il processo X_t il cui generatore infinitesimale è $(1/2)\Delta_g$, dove Δ_g è il laplaciano di Laplace–Beltrami. Questa definizione è intrinsecamente geometrica: il laplaciano è definito dalla metrica g , e rappresenta la divergenza del gradiente. Il moto browniano su una varietà è quindi la diffusione naturale determinata dalla geometria. La sua densità soddisfa l'equazione del calore $\partial_t p = (1/2)\Delta_g p$. La probabilità continua diventa una teoria della diffusione geometrica.

Per costruire il moto browniano su una varietà, è necessario introdurre il concetto di trasporto parallelo e di connessione di Levi-Civita. Il rumore infinitesimale dB_t vive nello spazio tangente $T_{\{X_t\}}M$, e deve essere trasportato lungo la traiettoria. La SDE su una varietà non può essere scritta come $dX_t = \sigma(X_t) dB_t$, perché dB_t non è un vettore nello spazio tangente. La formulazione corretta è $dX_t = \sum_i V_i(X_t) \circ dB_t^i$, dove V_i sono campi vettoriali su M e $\circ$ indica l'integrale di Stratonovich. Questa forma è necessaria perché l'integrale di Stratonovich rispetta la regola della catena classica, e quindi si comporta bene sotto cambiamenti di coordinate. L'integrale di Itô, invece, introduce termini di correzione che dipendono dalla metrica. La geometria stocastica richiede l'uso dell'integrale di Stratonovich per mantenere la coerenza geometrica.

La SDE su una varietà è quindi una dinamica che vive nello spazio tangente, e la sua evoluzione è determinata dalla connessione. Se si considera una SDE del tipo $dX_t = V_0(X_t) dt + \sum_i V_i(X_t) \circ dB_t^i$, il generatore infinitesimale è $\mathcal{L}f = V_0(f) + (1/2) \sum_i V_i(V_i(f))$. Se i campi V_i generano la struttura della varietà, il processo esplora la varietà in modo naturale. La diffusione è determinata dalla geometria dei campi vettoriali. La SDE su una varietà è quindi una teoria della dinamica stocastica immersa nella geometria differenziale.

La geometria stocastica è una teoria della curvatura. La curvatura della varietà influenza profondamente la diffusione. Se la curvatura è positiva, il moto browniano tende a concentrarsi; se la curvatura è negativa, tende a divergere. La curvatura di Ricci appare naturalmente nella teoria delle SDE su varietà. Il teorema di Bochner afferma che la curvatura di Ricci controlla la dissipazione dell'energia del moto browniano. La curvatura influenza la velocità di mescolanza, la ricorrenza, la transienza, la stabilità. La probabilità continua diventa una teoria della curvatura stocastica.

La geometria stocastica è una teoria della misura. La misura invariante del moto browniano su una varietà compatta è la misura di volume indotta dalla metrica g . Questa misura è la soluzione stazionaria dell'equazione del calore. La diffusione esplora la varietà in modo uniforme rispetto alla misura di volume. La misura invariante è quindi una teoria della geometria integrale.

La geometria stocastica è una teoria dei flussi stocastici. Una SDE su una varietà genera un flusso stocastico $\varphi_{\{s,t\}} : M \rightarrow M$, che trasporta punti lungo traiettorie stocastiche. Questo flusso non è invertibile in senso classico, ma è invertibile in senso probabilistico. La teoria dei flussi stocastici è una teoria della dinamica stocastica globale. La geometria stocastica diventa una teoria della trasformazione di misure su varietà.

La geometria stocastica è una teoria della parallelizzazione del rumore. Il rumore infinitesimale dB_t deve essere trasportato lungo la traiettoria tramite il trasporto parallelo. Questo trasporto è determinato dalla connessione di Levi-Civita. La SDE su una varietà è quindi una teoria del rumore trasportato lungo la geometria. La probabilità continua diventa una teoria della connessione stocastica.

La geometria stocastica è anche una teoria della termodinamica geometrica. Se la SDE è reversibile, la misura invariante è $\pi(x) \propto e^{-V(x)}$, dove V è un potenziale definito sulla varietà. La dinamica è un moto browniano in un paesaggio energetico curvo. La geometria e il potenziale determinano la struttura globale della misura invariante. La probabilità continua diventa una teoria della termodinamica su varietà.

Infine, la geometria stocastica è una teoria della diffusione su spazi complessi. Le varietà possono essere spazi di configurazioni, gruppi di Lie, superfici curve, spazi di forme, spazi di probabilità. La SDE su una varietà è una teoria della dinamica stocastica su spazi complessi. La probabilità continua diventa una teoria della diffusione geometrica.

La geometria stocastica e le SDE su varietà riemanniane rappresentano quindi uno dei vertici concettuali della probabilità continua. Esse unificano la dinamica stocastica, la geometria differenziale, la teoria delle misure, la termodinamica, la curvatura, la connessione, la diffusione. Sono la chiave per comprendere la dinamica stocastica su spazi curvi, e tutto ciò che seguirà — teoria dei campi stocastici, misure gaussiane su spazi funzionali, teoria di Malliavin — si innesta su questa struttura.

Campi gaussiani, misure su spazi funzionali e teoria di Malliavin

La teoria dei campi gaussiani rappresenta il punto in cui la probabilità continua si estende oltre i processi indicizzati dal tempo, abbracciando spazi di funzioni, superfici, distribuzioni e oggetti infinitamente dimensionali. Un campo gaussiano non è un processo stocastico nel senso classico, ma una famiglia $\{X_\varphi : \varphi \in \mathcal{F}\}$ di variabili aleatorie indicizzate da funzioni test, distribuzioni o punti di uno spazio, tale che ogni combinazione lineare finita è gaussiana. Questa definizione, apparentemente astratta, è la chiave per comprendere fenomeni che vanno dalla fisica statistica alla geometria aleatoria, dalla teoria quantistica dei campi alla probabilità su varietà, dalla teoria del rumore bianco alla regolarità delle soluzioni di equazioni stocastiche alle derivate parziali. La probabilità continua diventa una teoria della misura su spazi funzionali, una teoria della struttura gaussiana in dimensione infinita, una teoria della regolarità e della derivabilità stocastica.

Il punto di partenza è il rumore bianco gaussiano. In $\mathbb{R}$, il moto browniano B_t è un integrale del rumore bianco W , nel senso che $B_t = \int_0^t W_s ds$. Ma W non è una funzione: è una distribuzione, un oggetto che non può essere valutato punto per punto. Il rumore bianco è definito come un campo gaussiano tale che $E[W(\varphi) W(\psi)] = \int \varphi(x) \psi(x) dx$. Questa identità definisce la struttura di covarianza del rumore bianco. Il rumore bianco è quindi una misura gaussiana su uno spazio di distribuzioni. La probabilità continua diventa una teoria della misura su spazi di distribuzioni, non su spazi di funzioni.

Il moto browniano stesso può essere visto come un campo gaussiano indicizzato da funzioni test: $B(\varphi) = \int \varphi(t) dB_t$. Questa rappresentazione mostra che il moto browniano è un oggetto lineare rispetto alle funzioni test, e che la sua struttura è determinata dalla covarianza $E[B(\varphi) B(\psi)] = \iint \varphi(s) \psi(t) \min(s,t) ds dt$. La teoria dei campi gaussiani permette di trattare il moto browniano come un oggetto distribuzionale, non come una funzione. Questa prospettiva è fondamentale per comprendere la teoria di Malliavin.

La teoria di Malliavin nasce dall'esigenza di definire una derivata su spazi di funzioni aleatorie. Se X è una variabile aleatoria definita come un funzionale del moto browniano, $X = F(B)$, la teoria di Malliavin definisce una derivata $D_t X$ che misura la sensibilità di X rispetto al valore infinitesimale del rumore in t . Questa derivata non è una derivata classica: è una derivata in uno spazio di Hilbert infinito-dimensionale, l'Hilbert di Cameron–Martin. La teoria di Malliavin è quindi una teoria della derivabilità stocastica, una teoria del calcolo differenziale su spazi di cammini.

La derivata di Malliavin $D_t X$ è definita inizialmente per funzionali cilindrici, cioè per variabili aleatorie della forma $F(B) = f(B_{\{t_1\}}, \dots, B_{\{t_n\}})$, e poi estesa per chiusura. La derivata misura come F cambia se si perturba il cammino del moto browniano in modo infinitesimale lungo una direzione h del Cameron–Martin. Questa direzione h è una funzione assolutamente continua con derivata quadrato-integrabile. La teoria di Malliavin è quindi una teoria della sensibilità del funzionale rispetto a perturbazioni lisce del cammino, non rispetto a perturbazioni del rumore bianco. La distinzione è sottile ma fondamentale: il rumore bianco è troppo irregolare per essere perturbato direttamente; il cammino del moto browniano può essere perturbato lungo direzioni lisce.

La teoria di Malliavin permette di definire l'integrale divergente $\delta(u)$, che è il duale della derivata. Se D è la derivata, δ è la sua adjoint. L'integrale divergente generalizza l'integrale di Itô: se u_t è adattato, $\delta(u)$ coincide con $\int u_t dB_t$; se u_t non è adattato, $\delta(u)$ è un integrale anticipativo. La teoria di Malliavin è quindi una teoria dell'integrazione anticipativa, una teoria che estende l'integrale di Itô oltre il dominio adattato. Questa estensione è fondamentale per la teoria dei campi, per le equazioni stocastiche non adattate, per la fisica matematica.

La teoria di Malliavin è una teoria della regolarità delle distribuzioni di variabili aleatorie. Il teorema fondamentale afferma che, se una variabile aleatoria X ha una derivata di Malliavin non degenerata, allora X ha una densità liscia rispetto alla misura di Lebesgue. Questa proprietà è straordinaria: la derivata stocastica determina la regolarità della distribuzione. La teoria di Malliavin è quindi una teoria della liscezza delle densità, una teoria della regolarità stocastica.

La teoria di Malliavin è anche una teoria della geometria stocastica. La derivata $D_t X$ può essere vista come un vettore nello spazio tangente stocastico. La matrice di Malliavin $\gamma = \int D_t X (D_t X)^T dt$ è una matrice di covarianza stocastica. La non degenerazione di γ è la condizione di Hörmander per la regolarità delle densità. La teoria di Malliavin è quindi una teoria della geometria stocastica, una teoria della struttura infinitesimale dei processi.

La teoria dei campi gaussiani e la teoria di Malliavin si incontrano nella teoria delle misure gaussiane su spazi di Hilbert. Una misura gaussiana su uno spazio di Hilbert H è determinata da un operatore di covarianza Q . La struttura di Q determina la regolarità del campo. Se Q è compatto, il campo è irregolare; se Q è l'identità, il campo è rumore bianco. La teoria delle misure gaussiane è quindi una teoria della regolarità dei campi.

La teoria dei campi gaussiani è anche una teoria della geometria aleatoria. Campi come il campo gaussiano libero, il campo logaritmico, il campo di Ginzburg–Landau, sono oggetti centrali nella fisica statistica e nella teoria quantistica dei campi. Le loro proprietà geometriche — linee di livello, massimi, minimi, nodi — sono oggetti di studio profondi. La probabilità continua diventa una teoria della geometria aleatoria.

Infine, la teoria dei campi gaussiani e la teoria di Malliavin sono la base per la teoria moderna delle equazioni stocastiche alle derivate parziali. Le SPDE, come l'equazione del calore stocastica o l'equazione di KPZ, richiedono una teoria della regolarità stocastica che combina campi gaussiani, distribuzioni, derivata di Malliavin, teoria della rinormalizzazione. La probabilità continua diventa una teoria della dinamica stocastica in dimensione infinita.

La teoria dei campi gaussiani, delle misure su spazi funzionali e della derivata di Malliavin rappresenta quindi uno dei vertici concettuali della probabilità moderna. Essa unifica la teoria del rumore bianco, la teoria del moto browniano, la teoria della regolarità, la geometria stocastica, la teoria delle misure gaussiane, la teoria delle SPDE. È la teoria che permette di comprendere la struttura infinitesimale dei processi stocastici in dimensione infinita, e costituisce il ponte naturale verso il capitolo successivo, dedicato alla probabilità su spazi di distribuzioni e alla teoria dei campi stocastici non gaussiani.

Probabilità su spazi di distribuzioni e campi stocastici non gaussiani

La probabilità su spazi di distribuzioni rappresenta il punto in cui la teoria dei processi stocastici abbandona definitivamente il dominio delle funzioni e si immerge in quello, più vasto e più sottile, delle distribuzioni generalizzate. In questo contesto, un "campo stocastico" non è più una funzione $X(x)$ definita punto per punto, ma un funzionale $X(\varphi)$ definito per ogni funzione test φ , lineare e

continuo in φ , e aleatorio nella sua valutazione. Questa generalizzazione è necessaria perché molti oggetti fondamentali della probabilità moderna — il rumore bianco, i campi quantistici, le soluzioni di equazioni stocastiche alle derivate parziali — non sono funzioni, ma distribuzioni. La probabilità continua diventa una teoria della misura su spazi funzionali che non sono spazi di funzioni, ma spazi di distribuzioni, dotati di topologie deboli, di strutture di dualità, di proprietà di regolarità estremamente sottili.

Il punto di partenza è il rumore bianco gaussiano, che abbiamo già incontrato come campo gaussiano $W(\varphi)$ con covarianza $E[W(\varphi) W(\psi)] = \int \varphi(x) \psi(x) dx$. Questo oggetto non può essere rappresentato come una funzione $W(x)$: non esiste alcuna funzione che, integrata contro φ , produca $W(\varphi)$. Il rumore bianco è una distribuzione aleatoria. La sua natura distribuzionale è ciò che permette di definire equazioni stocastiche alle derivate parziali come $\partial_t u = \Delta u + \xi$, dove ξ è un rumore bianco spaziotemporale. Questa equazione non ha senso classico, perché ξ non è una funzione; ha senso solo come equazione in distribuzioni. La probabilità su spazi di distribuzioni è quindi una teoria della dinamica stocastica in dimensione infinita.

Ma il rumore bianco è solo il primo esempio. Molti campi stocastici non gaussiani, come il campo di Poisson, il campo di Lévy, il campo logaritmico, il campo di Ginzburg–Landau, sono distribuzioni aleatorie. La loro struttura non è determinata da una covarianza, ma da una misura di Lévy, da un potenziale, da una struttura di rinormalizzazione. La probabilità su spazi di distribuzioni è quindi una teoria della struttura infinitesimale dei campi non gaussiani.

Il campo di Poisson, ad esempio, è definito come $X(\varphi) = \sum \varphi(x_i)$, dove $\{x_i\}$ è un processo di Poisson. Questo campo non è gaussiano: ha salti, discontinuità, irregolarità. La sua struttura è determinata dalla misura di intensità del processo di Poisson. Il campo di Poisson è un esempio di campo di Lévy, cioè di campo con incrementi indipendenti ma non gaussiani. La teoria dei campi di Lévy è una teoria della discontinuità stocastica, una teoria della struttura a salti, una teoria della misura su spazi di distribuzioni con discontinuità.

Il campo logaritmico, invece, è un campo gaussiano con covarianza $E[X(x) X(y)] = -\log|x-y|$. Questo campo è estremamente irregolare: non è una funzione, non è una distribuzione classica, è un oggetto che vive al limite della teoria delle distribuzioni. La sua struttura è fondamentale nella teoria dei polimeri, nella teoria dei vortici, nella teoria dei campi conformi. La probabilità su spazi di distribuzioni è quindi una teoria della regolarità estrema.

Il campo di Ginzburg–Landau è un campo non gaussiano definito da una misura di Gibbs $\mu(dX) \propto \exp(-\int (|\nabla X|^2 + X^4) dx) dX$. Questa misura non è gaussiana, non è lineare, non è determinata da una covarianza. La sua struttura è determinata da un potenziale quartico. La teoria dei campi di Ginzburg–Landau è una teoria della non linearità stocastica, una teoria della misura su spazi funzionali con interazioni. La probabilità su spazi di distribuzioni diventa una teoria della fisica statistica.

La difficoltà fondamentale della probabilità su spazi di distribuzioni è che molti campi non gaussiani non sono ben definiti senza rinormalizzazione. Il prodotto di distribuzioni non è definito in generale, e molte equazioni stocastiche alle derivate parziali richiedono di definire prodotti come X^2 quando X è una distribuzione. La teoria della rinormalizzazione, sviluppata nella fisica quantistica e poi formalizzata nella probabilità, permette di definire questi prodotti sottraendo infiniti, regolarizzando, rinormalizzando. La probabilità su spazi di distribuzioni è quindi una teoria della rinormalizzazione stocastica.

La teoria moderna delle SPDE, come l'equazione di KPZ o l'equazione di Φ^4 , richiede una combinazione di teoria dei campi gaussiani, teoria di Malliavin, teoria della rinormalizzazione, teoria

delle distribuzioni. L'equazione di KPZ, ad esempio, $\partial_t h = (1/2)(\partial_x h)^2 + (1/2) \partial_{xx} h + \xi$, richiede di definire $(\partial_x h)^2$ quando h è una distribuzione. La soluzione è la rinormalizzazione: si sottrae un infinito, si definisce un prodotto rinormalizzato, si ottiene una soluzione ben definita. La probabilità su spazi di distribuzioni diventa una teoria della dinamica stocastica rinormalizzata.

La probabilità su spazi di distribuzioni è anche una teoria della misura infinita-dimensionale. Le misure gaussiane su spazi di Hilbert sono ben comprese; le misure non gaussiane sono molto più difficili. La loro struttura dipende da potenziali, da operatori differenziali, da condizioni di coercività. La probabilità su spazi di distribuzioni è quindi una teoria della misura in dimensione infinita.

Infine, la probabilità su spazi di distribuzioni è una teoria della geometria aleatoria. I campi stocastici definiscono superfici aleatorie, curve aleatorie, strutture geometriche aleatorie. La teoria delle linee di livello, dei nodi, dei massimi, dei minimi, è una teoria della geometria aleatoria. La probabilità continua diventa una teoria della forma aleatoria.

La probabilità su spazi di distribuzioni e i campi stocastici non gaussiani rappresentano quindi uno dei vertici concettuali della probabilità moderna. Essi unificano la teoria dei campi gaussiani, la teoria dei campi di Lévy, la teoria della rinormalizzazione, la teoria delle SPDE, la teoria della misura in dimensione infinita, la geometria aleatoria. Sono la chiave per comprendere la struttura infinitesimale dei processi stocastici in dimensione infinita, e costituiscono il ponte naturale verso la sintesi finale del Volume III, in cui la probabilità continua verrà ricomposta come una teoria unificata della dinamica, della misura, della geometria e dell'informazione.

Sintesi: la probabilità come dinamica, geometria e misura in dimensione infinita

La probabilità continua, dopo aver attraversato i processi stocastici, il moto browniano, le equazioni differenziali stocastiche, la dinamica delle densità, l'ergodicità, il filtraggio, la geometria stocastica e i campi in dimensione infinita, si rivela come una teoria unificata della dinamica, della misura e della geometria. Non è più una teoria di variabili aleatorie isolate, né una teoria di distribuzioni statiche, né una teoria di trasformazioni discrete. È una teoria della forma del caso nello spazio e nel tempo, una teoria della struttura infinitesimale dei cammini, una teoria della diffusione su varietà, una teoria della misura su spazi funzionali, una teoria della regolarità e della rinormalizzazione. La probabilità continua è una teoria della complessità infinitesimale.

Il punto di partenza è la misura sui cammini. Un processo stocastico non è una sequenza di variabili aleatorie, ma una misura μ su uno spazio di funzioni. Questa misura è l'oggetto fondamentale. Le variabili X_t non sono che proiezioni di questa misura. La probabilità continua è quindi una teoria della misura su spazi di funzioni, non su spazi di stati. Il moto browniano è la misura fondamentale su $C([0, \infty), \mathbb{R})$, e la sua struttura infinitesimale — la variazione quadratica, la continuità senza derivabilità, la natura frattale — è la firma del rumore gaussiano. L'integrale di Itô è la struttura analitica che permette di manipolare questa misura, di definire integrali rispetto al rumore, di costruire dinamiche stocastiche.

Le equazioni differenziali stocastiche sono la dinamica naturale su spazi di cammini. Una SDE non è un'equazione differenziale nel senso classico, ma una trasformazione della misura di Wiener. La soluzione X_t è un funzionale del cammino B_t , e la sua distribuzione è l'immagine della misura di Wiener tramite questa trasformazione. La SDE è quindi una teoria della trasformazione di misure su spazi di funzioni. La formula di Itô è la regola della catena che governa questa trasformazione, e

l'equazione di Fokker–Planck è la controparte densitaria della SDE. La probabilità continua è una teoria unificata della dinamica dei cammini e della dinamica delle densità.

La teoria ergodica stocastica è la teoria del comportamento asintotico delle SDE. La misura invariante π è il punto fisso della dinamica densitaria, l'equilibrio tra deriva e diffusione, la struttura globale determinata dalla dinamica locale. La convergenza verso π è la manifestazione della stabilità stocastica. La ricorrenza, la transienza, la mescolanza, la metastabilità, sono proprietà globali della dinamica. La probabilità continua è una teoria della stabilità e dell'instabilità in presenza di rumore.

La teoria dei filtri stocastici è la teoria dell'informazione continua. La distribuzione condizionata $\pi_t(dx) = \mu(X_t \in dx \mid \mathcal{F}_t)$ è una variabile aleatoria a valori nello spazio delle misure. La sua evoluzione è una SDE nello spazio delle misure. La teoria dei filtri è quindi una teoria della dinamica della conoscenza, una teoria della credenza che evolve sotto l'effetto simultaneo della dinamica e dell'osservazione. L'innovazione è il rumore bianco che rappresenta l'informazione nuova. La probabilità continua è una teoria dell'informazione infinitesimale.

La geometria stocastica è la teoria della diffusione su varietà. Il moto browniano su una varietà è la diffusione determinata dal laplaciano di Laplace–Beltrami. La SDE su una varietà è una dinamica immersa nella geometria differenziale. La curvatura influenza la diffusione, la ricorrenza, la mescolanza. La misura invariante è la misura di volume o una misura di Gibbs. La probabilità continua è una teoria della diffusione geometrica.

La teoria dei campi gaussiani è la teoria della misura su spazi di distribuzioni. Il rumore bianco è una distribuzione aleatoria, non una funzione. I campi gaussiani sono misure su spazi di distribuzioni, e la loro struttura è determinata da operatori di covarianza. La teoria di Malliavin è la teoria della derivabilità stocastica, una teoria del calcolo differenziale su spazi di cammini. La regolarità delle densità è determinata dalla non degenerazione della matrice di Malliavin. La probabilità continua è una teoria della derivabilità in dimensione infinita.

I campi non gaussiani, come i campi di Lévy, i campi logaritmici, i campi di Ginzburg–Landau, sono misure su spazi di distribuzioni con interazioni. La loro struttura richiede la rinormalizzazione. Le SPDE non lineari richiedono la teoria della regolarità stocastica, la teoria della rinormalizzazione, la teoria delle distribuzioni. La probabilità continua è una teoria della dinamica stocastica rinormalizzata.

La sintesi finale è che la probabilità continua è una teoria unificata della dinamica, della misura e della geometria in dimensione infinita. È una teoria della forma del caso, della struttura infinitesimale dei cammini, della diffusione su varietà, della misura su spazi funzionali, della regolarità stocastica, della rinormalizzazione. È una teoria che unifica la dinamica locale e la struttura globale, la misura e la densità, la conoscenza e l'osservazione, la geometria e il rumore, la funzione e la distribuzione. È una teoria che prepara naturalmente il terreno per il Volume IV, in cui la probabilità verrà integrata con la logica, la computazione, la semantica, la teoria dei programmi, la teoria delle categorie, la teoria dell'informazione e la teoria dei sistemi complessi.